

REMARKS

The specification has been amended to capitalize trademarks and remove reference to embedded hyperlinks.

Applicants have cancelled Claims 1-3, 7-10 and 15 without prejudice to, or disclaimer of, the subject matter contained therein. Applicants maintain that the cancellation of a claim makes no admission as to its patentability and reserve the right to pursue the subject matter of the cancelled claim in this or any other patent application.

Applicants have amended Claims 4, 5, 6 and 14 to delete elements (a)-(d). Claims 4 and 5 are amended to include the limitation "wherein said nucleic acid is more highly expressed in esophageal tumor and kidney tumor tissue compared to normal esophageal and normal kidney tissue." Claims 11 and 12 are amended to remove informalities. Claim 14 is amended to indicate that the isolated nucleic acid hybridizes under stringent conditions, and recites the stringent conditions. Claim 14 also is amended to include "or a complement thereof" to amended elements (a)-(c), and the following text "wherein said isolated nucleic acid molecule is suitable for use as a PCR primer, or probe; and wherein said isolated nucleic acid is at least about 450 nucleotides in length." Claim 16 is amended to read "at least about 500 nucleotides in length." Claim 17 is amended to depend from Claim 4. Claim 19 is amended to indicate that the cell is an isolated cell. New Claims 21-31 have been added.

Applicants submit that no new matter has been added by the amendments, and that support for the amendments can be found throughout the specification. For example, support for the amendment to Claims 4 and 5 regarding differential expression in esophageal tumor and kidney tumor can be found in Example 18 beginning at paragraph [0529], as well as paragraph [0336] of the specification. Support for the amendments to Claim 14 can be found, for example, at paragraphs [0012], [0317], and [0327] of the specification. Support for the amendment to Claim 16 and new Claims 21-25 can be found, for example, at paragraph [0012]. Support for new Claims 26-31 can be found, for example, in the claims as originally filed, and paragraphs [0227] and [0317].

The rejections of the presently pending claims are respectfully traversed. Claims 4-6, 11-14, and 16-31 are presented for examination.

Appl. No. : 10/063,713
Filed : May 8, 2002

Correction of Inventorship under 37 CFR §1.48(b)

Applicants request that several inventors be deleted, as these inventors' inventions are no longer being claimed in the present application as a result of prosecution. The fee as set forth in § 1.17(i) is submitted herewith.

Priority Determination:

As an initial matter, the PTO issued the instant Office Action assuming that the earliest priority is the instant filing date, May 8, 2002. The PTO argued that the instant application and priority application Serial No. 10/006,867 do not meet the requirements of 35 U.S.C. § 112, first paragraph. However, for the reasons set forth below, the instant application and the priority application do meet the requirements of 35 U.S.C. § 112, first paragraph, and therefore, are entitled to an earlier priority date.

Applicants have previously listed the priority information for the instant application in a Preliminary Amendment mailed September 5, 2002. The preliminary amendment states that the instant "application is a continuation of, and claims priority under 35 U.S.C. § 120 to, US Application 10/006867 filed 12/6/2001, which is a continuation of, and claims priority under 35 U.S.C. § 120 to, PCT Application PCT/US00/23328 filed 8/24/2000, which is a continuation-in-part of, and claims priority under 35 U.S.C. § 120 to, US Application 09/403297 filed 10/18/1999, now abandoned, which is the National Stage filed under 35 U.S.C. § 371 of PCT Application PCT/US99/20111 filed 9/1/1999, which claims priority under 35 U.S.C. § 119 to U.S. Provisional Application 60/105881 filed 10/27/1998."

The sequences of SEQ ID NOs: 81 and 82 were first disclosed in U.S. Provisional Application 60/105,881 filed 10/27/1998 as SEQ ID NO:1 and 2 and in Figures 1 and 2. These same sequences were disclosed in PCT/US99/20111 and in 09/403,297 as SEQ ID NO:141 and 142, Figures 85 and 86. The data in Example 18 (Tumor Versus Normal Differential Tissue Expression Distribution), relied on in part for the utility of the claimed nucleic acids, were first disclosed in PCT Application PCT/US00/23328 filed 8/24/2000, on page 93, line 3, through page 96, line 35. Thus, Applicants maintain that the present application is fully entitled to the benefit of at least the priority date of August 24, 2000.

Appl. No. : 10/063,713
Filed : May 8, 2002

Rejection under 35 U.S.C. §101 – Utility

The PTO has rejected Claims 1-20 as lacking a specific, substantial, and credible utility. The PTO argues that utilities asserted in the specification are not specific and substantial or well established. According to the PTO, “[t]he encoding nucleic acid cannot derive a utility from the encoded polypeptide because there is neither a known physiological or clinical significance of the [encoded] polypeptide, and the prior art does not support a very close structural relationship to a well described (structurally and functionally) family of known proteins.” Office Action at 2.

The PTO cites Hu *et al.* (J. Proteome Res., 2(4):405-12 (2003)) and Wu *et al.* (Gene 311:105-110 (2003)) to support its assertion that the literature cautions against drawing conclusions based on small changes in transcript expression levels between normal and cancerous tissue, and that upregulation of a gene does not necessarily indicate that the tumor is vascularized.

One of the asserted utilities for the claimed nucleic acids is use as a diagnostic tool, as well as therapeutically as a target for treatment, based on the data that PRO1557 cDNA is more highly expressed in esophageal tumor and kidney tumor as compared to normal esophagus and normal kidney tissue, respectively. The PTO recognizes this as a “possible utility,” however, the PTO asserts that there is no guidance on how to use this information, that no levels are disclosed, and that the information is too sparse to allow the encoding polynucleotide to be used as a diagnostic marker for esophageal or kidney tumor.

The PTO also argues that even if the polynucleotide has utility as a tumor marker, there is no such utility for the polypeptide because there is no reason to suspect that there is an alteration in the amount of the polypeptide in normal esophagus and kidney tissue compared to esophageal and kidney tumor tissue. For the above reasons, the PTO asserts that there is no substantial and specific utility for the nucleic acid of SEQ ID NO:81.

Applicants respectfully disagree and submit that for the reasons stated below, the claimed nucleic acids have a credible, substantial, and specific utility.

Utility – Legal Standard

According to the Utility Examination Guidelines (“Utility Guidelines”), 66 Fed. Reg. 1092 (2001) an invention complies with the utility requirement of 35 U.S.C. § 101, if it has at least one asserted “specific, substantial, and credible utility” or a “well-established utility.”

Appl. No. : 10/063,713
Filed : May 8, 2002

Under the Utility Guidelines, a utility is “specific” when it is particular to the subject matter claimed. For example, it is generally not enough to state that a nucleic acid is useful as a diagnostic tool without also identifying the condition that is to be diagnosed.

The requirement of “substantial utility” defines a “real world” use, and derives from the Supreme Court’s holding in *Brenner v. Manson*, 383 U.S. 519, 534 (1966) stating that “The basic *quid pro quo* contemplated by the Constitution and the Congress for granting a patent monopoly is the benefit derived by the public from an invention with substantial utility.” In explaining the “substantial utility” standard, M.P.E.P. § 2107.01 cautions, however, that Office personnel must be careful not to interpret the phrase “immediate benefit to the public” or similar formulations used in certain court decisions to mean that products or services based on the claimed invention must be “currently available” to the public in order to satisfy the utility requirement. “Rather, *any reasonable use that an applicant has identified for the invention that can be viewed as providing a public benefit should be accepted as sufficient*, at least with regard to defining a ‘substantial’ utility.” (M.P.E.P. § 2107.01, emphasis added).

The mere consideration that further experimentation might be performed to more fully develop the claimed subject matter does not support a finding of lack of utility. M.P.E.P. § 2107.01 III cites *In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995) in stating that “Usefulness in patent law ... necessarily includes the expectation of further research and development. The stage at which an invention in this field becomes useful is well before it is ready to be administered to humans.” Further, “[T]o violate § 101 the claimed device must be totally incapable of achieving a useful result” *Juicy Whip Inc. v. Orange Bang Inc.*, 51 USPQ2d 1700 (Fed. Cir. 1999), citing *Brooktree Corp. v. Advanced Micro Devices, Inc.*, 977 F.2d 1555, 1571 (Fed.Cir.1992).

Indeed, the Guidelines for Examination of Applications for Compliance With the Utility Requirement, set forth in M.P.E.P. § 2107 II(B)(1) gives the following instruction to patent examiners: “If the applicant has asserted that the claimed invention is useful for any particular practical purpose ... and the assertion would be considered credible by a person of ordinary skill in the art, do not impose a rejection based on lack of utility.”

Appl. No. : 10/063,713
Filed : May 8, 2002

Utility need NOT be Proved to a Statistical Certainty – a Reasonable Correlation between the Evidence and the Asserted Utility is Sufficient

An Applicant's assertion of utility creates a presumption of utility that will be sufficient to satisfy the utility requirement of 35 U.S.C. § 101, “unless there is a reason for one skilled in the art to question the objective truth of the statement of utility or its scope.” *In re Langer*, 503 F.2d 1380, 1391, 183 USPQ 288, 297 (CCPA 1974). *See, also In re Jolles*, 628 F.2d 1322, 206 USPQ 885 (CCPA 1980); *In re Irons*, 340 F.2d 974, 144 USPQ 351 (1965); *In re Sichert*, 566 F.2d 1154, 1159, 196 USPQ 209, 212-13 (CCPA 1977). Compliance with 35 U.S.C. § 101 is a question of fact. *Raytheon v. Roper*, 724 F.2d 951, 956, 220 USPQ 592, 596 (Fed. Cir. 1983) cert. denied, 469 US 835 (1984). The evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the evidence, or “more likely than not” standard. *In re Oetiker*, 977 F.2d 1443, 1445, 24 USPQ2d 1443, 1444 (Fed. Cir. 1992). This is stated explicitly in the M.P.E.P.:

[T]he applicant does not have to provide evidence sufficient to establish that an asserted utility is true “beyond a reasonable doubt.” **Nor must the applicant provide evidence such that it establishes an asserted utility as a matter of statistical certainty.** Instead, evidence will be sufficient if, considered as a whole, it leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true. M.P.E.P. at § 2107.02, part VII (2004) (underline emphasis in original, bold emphasis added, internal citations omitted).

The PTO has the initial burden to offer evidence “that one of ordinary skill in the art would reasonably doubt the asserted utility.” *In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). Only then does the burden shift to the Applicant to provide rebuttal evidence. *Id.* As stated in the M.P.E.P., such rebuttal evidence does not need to absolutely prove that the asserted utility is real. Rather, the evidence only needs to be reasonably indicative of the asserted utility.

In *Fujikawa v. Wattanasin*, 93 F.3d 1559, 39 U.S.P.Q.2d 1895 (Fed. Cir. 1996), the Court of Appeals for the Federal Circuit upheld a PTO decision that *in vitro* testing of a novel pharmaceutical compound was sufficient to establish practical utility, stating the following rule:

[T]esting is often required to establish practical utility. But the test results **need not absolutely prove** that the compound is pharmacologically active. All that is required is that the tests be “*reasonably* indicative of the desired [pharmacological] response.” In other words, there must be **a sufficient correlation** between the tests and an asserted pharmacological activity so as to

Appl. No. : 10/063,713
Filed : May 8, 2002

convince those skilled in the art, **to a reasonable probability**, that the novel compound will exhibit the asserted pharmacological behavior.” *Fujikawa v. Wattanasin*, 93 F.3d 1559, 1564, 39 U.S.P.Q.2d 1895 (Fed. Cir. 1996) (internal citations omitted, bold emphasis added, italics in original).

While the *Fujikawa* case was in the context of utility for pharmaceutical compounds, the principals stated by the Court are applicable in the instant case where the asserted utility is for a therapeutic and diagnostic use – utility does not have to be established to an absolute certainty, rather, the evidence must convince a person of skill in the art “to a reasonable probability.” In addition, the evidence need not be direct, so long as there is a “sufficient correlation” between the tests performed and the asserted utility.

Thus, the legal standard for demonstrating utility is a relatively low hurdle. An Applicant need only provide evidence such that it is **more likely than not that a person of skill in the art would be convinced, to a reasonable probability, that the asserted utility is true.** The evidence need not be direct evidence, so long as there is a reasonable correlation between the evidence and the asserted utility. The Applicant **does not need to provide evidence such that it establishes an asserted utility as a matter of statistical certainty.**

Even assuming that the PTO has met its initial burden to offer evidence that one of ordinary skill in the art would reasonably doubt the truth of the asserted utility, Applicants assert that they have met their burden of providing rebuttal evidence such that it is more likely than not those skilled in the art, to a reasonable probability, would believe that the claimed invention is useful as a diagnostic tool for cancer.

Substantial Utility

Summary of Applicants’ Arguments and the PTO’s Response

In an attempt to clarify Applicants’ argument, Applicants offer a summary of their argument and the disputed issues involved. Applicants assert that the claimed nucleic acids have utility as diagnostic tools for cancer, particularly esophageal and kidney cancer. Applicants’ asserted utility rests on the following argument:

1. Applicants assert they have provided reliable evidence that mRNA for the PRO1557 polypeptide is expressed at least two-fold higher in esophageal tumor and kidney tumor compared to normal esophageal and kidney tissue, and therefore the claimed nucleic acids are

Appl. No. : 10/063,713
Filed : May 8, 2002

useful as diagnostic tools. Applicants are not asserting that the claimed nucleic acids will necessarily provide a definitive diagnosis of cancer, but rather that they are useful, alone or in combination with other diagnostic tools to assist in the diagnosis of certain cancers.

2. Applicants submit that it is not necessary to know what role the PRO1557 gene plays in cancer to use its differential expression as a diagnostic tool.

3. It is not required to prove that the PRO1557 polypeptide is also differentially expressed in certain tumors to establish the utility of the claimed nucleic acids.

Applicants understand the PTO to be making several arguments in response to Applicants' asserted utility:

1. The PTO has challenged the reliability of the evidence reported in Example 18, and states that it does not provide the expression levels, and that the information is too sparse to allow the encoding polynucleotide to be used as a diagnostic marker for tumors;

2. The PTO cites Hu *et al.* and Wu *et al.* for the assertion that the literature cautions against drawing conclusions based on small changes in transcript expression levels, and that upregulation of a gene does not necessarily indicate that the tumor is vascularized;

3 The PTO asserts that the nucleic acid cannot derive utility from the encoded polypeptide because there is no known physiological or clinical significance of the polypeptide and because there is no reason to think that there is alteration of encoded polypeptide in esophageal tumor or kidney tumor relative to normal esophageal or kidney tissue.

As detailed below, Applicants submit that the PTO has failed to meet its initial burden to offer evidence "that one of ordinary skill in the art would reasonably doubt the asserted utility." *In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). First, Applicants submit herewith a copy of a declaration of J. Christopher Grimaldi, (attached as Exhibit 1) which establishes the reliability of the data of Example 18. Second, the references provided by the PTO not contrary to Applicants' arguments and evidence, and therefore are not evidence to support the PTO's position. Third, Applicants submit that given the well-established correlation between a change in the level of mRNA with a corresponding change in the levels of the encoded protein, the PRO1557 protein is likely differentially expressed in certain tumors. However, utility for the pending claims does not rely on whether the encoded polypeptide is overexpressed, and as such whether or not increased levels of PRO1557 mRNA correlate with increased levels of PRO1557

Appl. No. : 10/063,713
Filed : May 8, 2002

protein is not presently an issue. Fourth, Applicants do not rely on the function of the encoded polypeptides for utility for the claimed nucleic acids.

Finally, even if the PTO has met its initial burden, Applicants have submitted enough rebuttal evidence such that it is **more likely than not** that a person of skill in the art would be convinced, **to a reasonable probability**, that the asserted utility is true. As stated above, Applicants' evidence need not be direct evidence, so long as there is a reasonable correlation between the evidence and the asserted utility. **The standard is not absolute or statistical certainty.**

Applicants have established that the Gene Encoding the PRO1557 Polypeptide is Differentially Expressed in Certain Cancers compared to Normal Tissue and is Useful as a Diagnostic Tool

Applicants first address the PTO's argument that the evidence of differential expression of the gene encoding the PRO1557 polypeptide in certain tumors compared to their normal counterparts is insufficient because the specification provides no information regarding values of the differences in transcript levels, and the disclosure of the specification is too sparse. Applicants also address the PTO's argument that the data do not establish a utility because the specification does not disclose any information on the level of expression, activity, or role of the PRO1557 polypeptide in cancer. Applicants submit that the gene expression data provided in Example 18 of the present application are sufficient to establish a specific and substantial utility for the claimed nucleic acids related to the gene encoding the PRO1557 polypeptide.

Applicants submit herewith a copy of a declaration of J. Christopher Grimaldi, an expert in the field of cancer biology, originally submitted in a related co-pending and co-owned patent application Serial No. 10/063,557 (attached as Exhibit 1). In paragraphs 6 and 7, Mr. Grimaldi explains that the semi-quantitative analysis employed to generate the data of Example 18 is sufficient to determine if a gene is over- or underexpressed in tumor cells compared to corresponding normal tissue. He states that any visually detectable difference seen between two samples is indicative of at least a two-fold difference in cDNA between the tumor tissue and the counterpart normal tissue. Thus, the results of Example 18 reflect at least a two-fold difference between normal and tumor samples.

He also states that the results of the gene expression studies indicate that the genes of interest "can be used to differentiate tumor from normal," thus establishing their reliability. He

Appl. No. : 10/063,713
Filed : May 8, 2002

explains that, contrary to the PTO's assertions, "The precise levels of gene expression are irrelevant; what matters is that there is a relative difference in expression between normal tissue and tumor tissue." (Paragraph 7). Thus, since it is the relative level of expression between normal tissue and suspected cancerous tissue that is important, the precise level of expression in normal tissue is irrelevant. Likewise, there is no need for quantitative data to compare the level of expression in normal and tumor tissue. As Mr. Grimaldi states, "If a difference is detected, this indicates that the gene and its corresponding polypeptide and antibodies against the polypeptide are useful for diagnostic purposes, to screen samples to differentiate between normal and tumor."

Applicants submit that a lack of known role for the gene encoding PRO1557 in cancer does not prevent its use as a diagnostic tool for cancer. Whether the differential expression of the gene encoding PRO1557 is a cause or result of the esophageal and kidney tumors is irrelevant to whether its differential expression can be used to assist in diagnosis of cancer – one does not need to know why the PRO1557 gene is differentially expressed, or what the consequence of the differential expression is, in order to exploit the differential expression to distinguish tumor from normal tissue.

The PTO has recognized that the utility of a nucleic acid does not depend on the function of the encoded gene product. The Utility Examination Guidelines published on January 5, 2001 state "In addition, the utility of a claimed DNA does not necessarily depend on the function of the encoded gene product. A claimed DNA may have a specific and substantial utility because, e.g. it hybridizes near a disease-associated gene or it has a gene regulating activity." (Federal Register, Volume 66, page 1095, Comment 14). While Applicants appreciate that actions taken in other applications are not binding on the PTO with respect to the present application, Applicants note that the PTO issues patents relating to nucleic acids which are useful for diagnosing particular conditions regardless of whether the nucleic acids are the causative agent for the condition. For example, polymorphisms which are indicative of a predisposition to a particular condition are patentable (*see, e.g.,* U.S. Patent No. 6,465,185, U.S. Patent No. 6,228,582, and U.S. Patent No. 6,162,604 submitted herewith as Exhibits 2-4), even though they may or may not cause the disease itself. Similarly, the present nucleic acids which are useful for determining whether an individual has cancer are useful regardless of whether or not they are the cause of the cancer.

Appl. No. : 10/063,713
Filed : May 8, 2002

The PTO relies on two references to support its assertion that the literature cautions researchers from drawing conclusions based on small changes in transcript expression levels between normal and cancerous tissue. The PTO cites Hu *et al.* (J. Proteome Res., 2(4):405-12 (2003)) for support for the conclusion that not all genes with increased expression in cancer have a known or published role in cancer. The PTO cites Wu *et al.* (Gene 311:105-110 (2003)) to support its assertion that upregulation of a gene does not necessarily indicate that the tumor is vascularized. Applicants respectfully submit that these references do not satisfy the PTO's burden to offer evidence that one of ordinary skill in the art would reasonably doubt the truth of the asserted utility.

In Hu, the researchers used an automated literature-mining tool to summarize and estimate the relative strengths of all human gene-disease relationships published on Medline. They then generated a microarray expression dataset comparing breast cancer and normal breast tissue. Using their data-mining tool, they looked for a correlation between the strength of the literature association between the gene and breast cancer, and the magnitude of the difference in expression level. They report that for genes displaying a 5-fold change or less in tumors compared to normal, there was no evidence of a correlation between altered gene expression and a *known* role in the disease. See Hu at 411. However, among genes with a 10-fold or more change in expression level, there was a strong correlation between expression level and a *published* role in the disease. *Id.* at 412. Importantly, Hu reports that the observed correlation was only found among estrogen receptor-positive tumors, not less-prevalent ER-negative tumors. *Id.*

The general findings of Hu are not surprising – one would expect that genes with the greatest change in expression in a disease would be the first targets of research, and therefore have the strongest known relationship to the disease as measured by the number of publications reporting a connection with the disease. The correlation reported in Hu only indicates that the greater the change in expression level, the more likely it is that there is a *published* or *known* role for the gene in the disease, as found by their automated literature-mining software. Thus, Hu's results merely reflect a bias in the literature toward studying the most prominent targets, and reflect nothing regarding the ability of a gene that is 2-fold or more differentially expressed in tumors to serve as a disease marker. Hu acknowledges the shortcomings of this method in explaining the disparity in Hu's findings for ER-negative versus ER-positive tumors: Hu

Appl. No. : 10/063,713
Filed : May 8, 2002

attributes the “bias in the literature” toward the more prevalent ER-positive tumors as the explanation for the lack of any correlation between number of publications and gene expression levels in less-prevalent (and, therefore, less studied) ER-negative tumors. *Id.* Because of this intrinsic bias, Hu’s methodology is unlikely to ever note a correlation of a disease with less differentially-expressed genes and their corresponding proteins, regardless of whether or not an actual relationship between the disease and less differentially-expressed genes exists. Accordingly, Hu’s methodology yields results that provide little or no information regarding biological significance of genes with less than 5-fold expression change in disease.

Applicants submit that a lack of known role for PRO1557 in cancer does not prevent its use as a diagnostic tool for cancer. There is a difference between use of a gene for distinguishing between tumor and normal tissue on the one hand, and establishing a role for the gene in cancer on the other. Genes with lower levels of change in expression may or may not be the most important genes in causing the disease, but the genes can still show a consistent and measurable change in expression. While such genes may or may not be good targets for further research, they can nonetheless be used as diagnostic tools. Thus, Hu does not refute the Applicants’ assertion that the PRO1557 gene can be used as a cancer diagnostic tool because it is differentially expressed in certain tumors.

The PTO also cites Wu *et al.* (Gene 311:105-110 (2003)) as support for the PTO’s assertion that upregulation of a gene does not necessarily indicate that the tumor is vascularized. Wu *et al.* identify a gene, BNF-1, as a putative extracellular matrix protein over-expressed in breast, lung and colon tumors, which were the only tumors tested. Wu found that in 3 out of 11 breast tumor samples, BNF-1 was up-regulated about 2-fold to 3-fold. Wu at 107. Wu found that BNF-1 was up-regulated about 2-fold to 3-fold in 2 out of 6 lung tumor samples. *Id.* at 109. Wu found that BNF-1 was up-regulated about 2-fold to about 4-fold in 1 out of 6 colon tumor samples. *Id.* The coding region of BNF-1 is identical to the coding region of SEQ ID NO:81. Thus, Wu demonstrates that a gene identical to that of Applicants claims is over-expressed by 2-fold to 4-fold in some tumor samples, and Wu concludes that this gene is up-regulated in tumors. Wu also notes that the commercially provided tumor samples did not indicate the vascular state of the source tumors. Wu states that the relationship between up-regulation of BNF-1 and tumor vascularization is not determined in this study.

Appl. No. : 10/063,713
Filed : May 8, 2002

The PTO cites Wu for the proposition that even if a gene is up-regulated in a tumor, this does not serve as an indication that the tumor is vascularized or malignant. Applicants do not disagree, but Applicants maintain that this position is irrelevant for the purposes of determining the utility of a gene. Applicants are not asserting that the claimed nucleic acids will necessarily provide a definitive diagnosis of vascularized cancer, but rather that the claimed nucleic acids are useful, alone or in combination with other diagnostic tools to assist in the diagnosis of certain cancers.

The PTO appears to require that, in order for gene asserted to be a tumor marker to have any utility whatsoever, there must be a demonstration that the gene in question is up-regulated in vascularized tumors. This position of the PTO is inconsistent with the analogous standard for therapeutic utility of a compound that “the mere identification of a pharmacological activity of a compound that is relevant to an asserted pharmacological use provides an ‘immediate benefit to the public’ and thus satisfies the utility requirement.” M.P.E.P. §2701.01 (emphasis original). Here, the mere identification of altered expression in tumors is relevant to diagnosis of tumors, and, therefore, provides an immediate benefit to the public. The position of the PTO is also inconsistent with the statements of Wu itself. Wu discusses that observations similar to that for BNF-1 have been made for other solid tumor oncogenes such as N-MYC. Wu at 109. Thus, Wu indicates that BNF-1 expression patterns is consistent with that observed for other oncogenes. Accordingly, far from serving as evidence of a lack of utility for the claimed nucleic acids, Wu supports their utility because BNF-1, which Wu states has expression patterns consistent with oncogenes, has the identical coding region to SEQ ID NO:81 recited in Applicants’ claims.

Moreover, Wu also serves as evidence contrary to the PTO’s position that changes in expression level below 5-fold are insufficient to supply utility (for which the PTO relies on the Hu reference). Wu reports only increases in BNF-1 expression of 2-fold to 4-fold. These results lead Wu to conclude that BNF-1 is up-regulated in tumors and that the expression pattern for BNF-1 is consistent with that of other solid tumor oncogenes. While Hu merely indicates that less differentially expressed genes are less-often the subjects of scientific publications, Wu asserts that the 2-fold to 4-fold overexpressed BNF-1 gene is consistent with other solid tumor oncogenes. Thus, the teachings of Wu toward the utility of BNF-1, and similarly up-regulated oncogenes in general, are more applicable to the question of the utility of Applicants’ claimed

Appl. No. : 10/063,713
Filed : May 8, 2002

nucleic acids than the teachings of Hu. Accordingly, the evidence presented by the PTO, as a whole, supports Applicants' assertion of utility of the claimed nucleic acids.

As stated above, the standard for utility is not absolute certainty, but rather whether one of skill in the art would be more likely than not to believe the asserted utility. Hu and Wu are not sufficient to prove that a person of skill in the art would consider it unlikely that a gene differentially expressed in certain tumors can be used as a diagnostic tool since neither reference teaches against this, and Wu supports Applicants' asserted utility of the claimed nucleic acids. Given the lack of support for the PTO's position, and the supporting evidence provided by the PTO and Applicants for Applicants' position, one of skill in the art would be more likely than not to believe that the claimed nucleic acids related to PRO1557 gene can be used as diagnostic tools for cancer, particularly esophageal and kidney cancer.

In conclusion, Applicants submit that the evidence reported in Example 18, combined with the first Grimaldi Declaration, establish that there is at least a two-fold difference in PRO1557 cDNA between esophageal tumor tissue and kidney tumor tissue, and normal esophageal tissue and normal kidney tissue, respectively. Therefore, it follows that expression levels of the PRO1557 gene can be used to distinguish esophageal tumor tissue from normal esophageal tissue and kidney tumor tissue from normal kidney tissue. The evidence offered by the PTO supports Applicants asserted utility without supporting any significant argument to the contrary. Applicants have therefore established a utility for the claimed nucleic acids as diagnostic tools for cancer, particularly esophageal and kidney tumors.

Applicants have established that the Accepted Understanding in the Art is that there is a Positive Correlation between mRNA Levels and the Level of Expression of the Encoded Protein

Applicants have asserted that there is a direct correlation between changes in the level of mRNA and changes in the level of expression of the corresponding protein. Because the claims have been amended such that the claimed nucleic acids are not defined by the sequence of the polypeptide they encode, the question of whether there is a correlation between changes in gene expression and changes in protein expression are not presently at issue. However, Applicants submit that they have established for the record that it is well-established in the art that a change in the level of mRNA for a particular protein, generally leads to a corresponding change in the level of the encoded protein. Given Applicants' evidence of differential expression of the

Appl. No. : 10/063,713
Filed : May 8, 2002

mRNA for the PRO1557 polypeptide in esophageal and kidney tissue, it is more likely than not that the PRO1557 polypeptide is also differentially expressed.

The PTO states that even if the claimed nucleic acid had utility, the “encoded polypeptide would have no such utility since there is no reason to suspect that there is alteration of polypeptide sequence or amount in esophageal or kidney tumor *versus* normal tissue.” Office action at 4 (emphasis original). No substantiating evidence is presented. This statement in the Office Action does not satisfy the PTO’s burden to offer evidence that one of ordinary skill in the art would reasonably doubt the truth of the asserted utility. As stated above, the standard for establishing a use for a claimed invention is not absolute or even statistical certainty, and thus a *necessary* correlation between mRNA levels and protein levels is not required.

The PTO cites no evidence that would cast any doubt on the Applicants assertion that in general, there is a positive correlation between changes in mRNA level and changes in the encoded protein level.

In further support of the assertion that changes in mRNA are positively correlated to changes in protein levels, Applicants submit herewith a copy of a second Declaration by J. Christopher Grimaldi, an expert in the field of cancer biology (attached as Exhibit 5). This declaration was submitted in connection with the related co-pending and co-owned application Serial No. 10/063,557. As stated in paragraph 5 of the declaration, “Those who work in this field are well aware that in the vast majority of cases, when a gene is over-expressed...the gene product or polypeptide will also be over-expressed.... This same principal applies to gene under-expression.” Further, “the detection of increased mRNA expression is expected to result in increased polypeptide expression, and the detection of decreased mRNA expression is expected to result in decreased polypeptide expression. The detection of increased or decreased polypeptide expression can be used for cancer diagnosis and treatment.” The references cited in the declaration and submitted herewith support this statement.

Applicants also submit herewith a copy of the declaration of Paul Polakis, Ph.D. (attached as Exhibit 6), an expert in the field of cancer biology, originally submitted in a related and co-owned patent application Serial No. 10/032,996. As stated in paragraph 6 of his declaration:

Based on my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above [showing a positive correlation between mRNA levels and encoded protein levels in the vast majority of cases] and my knowledge of the relevant scientific literature, it is my

Appl. No. : 10/063,713
Filed : May 8, 2002

considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, *it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.* (Emphasis added).

Dr. Polakis acknowledges that there are published cases where such a correlation does not exist, but states that it is his opinion, based on over 20 years of scientific research, that “such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.” (Polakis Declaration, paragraph 6).

The statements of Grimaldi and Polakis are supported by the teachings in Molecular Biology of the Cell, a leading textbook in the field (Bruce Alberts, *et al.*, Molecular Biology of the Cell (3rd ed. 1994) (submitted herewith as Exhibit 7) and (4th ed. 2002) (submitted herewith as Exhibit 8)). Figure 9-2 of Exhibit 7 shows the steps at which eukaryotic gene expression can be controlled. The first step depicted is transcriptional control. Exhibit 7 provides that “[f]or most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 9-2, only transcriptional control ensures that no superfluous intermediates are synthesized.” Exhibit 7 at 403 (emphasis added). In addition, the text states that “Although controls on the initiation of gene transcription are the predominant form of regulation for most genes, other controls can act later in the pathway from RNA to protein to modulate the amount of gene product that is made.” Exhibit 7 at 453 (emphasis added). Thus, as established in Exhibit 7, the predominant mechanism for regulating the amount of protein produced is by regulating transcription initiation.

In Exhibit 8, Figure 6-3 on page 302 illustrates the basic principle that there is a correlation between increased gene expression and increased protein expression. The accompanying text states that “a cell can change (or regulate) the expression of each of its genes according to the needs of the moment – *most obviously by controlling the production of its mRNA.*” Exhibit 8 at 302 (emphasis added). Similarly, Figure 6-90 on page 364 of Exhibit 8 illustrates the path from gene to protein. The accompanying text states that while potentially each step can be regulated by the cell, “the initiation of transcription is the most common point for a cell to regulate the expression of each of its genes.” Exhibit 8 at 364 (emphasis added).

Appl. No. : 10/063,713
Filed : May 8, 2002

This point is repeated on page 379, where the authors state that of all the possible points for regulating protein expression, “[f]or most genes transcriptional controls are paramount.” Exhibit 8 at 379 (emphasis added).

Further support for Applicants’ position can be found in the textbook, *Genes VI*, (Benjamin Lewin, *Genes VI* (1997)) (submitted herewith as Exhibit 9) which states “having acknowledged that control of gene expression can occur at multiple stages, and that production of RNA cannot inevitably be equated with production of protein, it is clear that the overwhelming majority of regulatory events occur at the initiation of transcription.” *Genes VI* at 847-848 (emphasis added).

Additional support is also found in Zhigang *et al.*, *World Journal of Surgical Oncology* 2:13, 2004, submitted herewith as Exhibit 10. Zhigang studied the expression of prostate stem cell antigen (PSCA) protein and mRNA to validate it as a potential molecular target for diagnosis and treatment of human prostate cancer. The data showed “a high degree of correlation between PSCA protein and mRNA expression.” Exhibit 10 at 4. Of the samples tested, 81 out of 87 showed a high degree of correlation between mRNA expression and protein expression. The authors conclude that “it is demonstrated that PSCA protein and mRNA overexpressed in human prostate cancer, and that the increased protein level of PSCA was resulted from the upregulated transcription of its mRNA.” Exhibit 10 at 6. Even though the correlation between mRNA expression and protein expression occurred in 93% of the samples tested, not 100%, the authors state that “PSCA may be a promising molecular marker for the clinical prognosis of human Pca and a valuable target for diagnosis and therapy of this tumor.” Exhibit 10 at 7.

Further, Meric *et al.*, *Molecular Cancer Therapeutics*, vol. 1, 971-979 (2002), submitted herewith as Exhibit 11, states the following:

The **fundamental principle** of molecular therapeutics in cancer is to exploit the differences in gene expression between cancer cells and normal cells...[M]ost efforts have concentrated on identifying differences in gene expression at the level of mRNA, which can be attributable to either DNA amplification or to differences in transcription. Meric *et al.* at 971 (emphasis added).

Those of skill in the art would not be focusing on differences in gene expression between cancer cells and normal cells if there were no correlation between gene expression and protein expression.

Appl. No. : 10/063,713
Filed : May 8, 2002

As discussed above, whether or not increased levels of PRO1557 mRNA correlate with increased levels of PRO1557 protein is not presently an issue. However, Applicants submit together, the declarations of Grimaldi and Polakis, the accompanying references, and the excerpts and references provided above all establish that the accepted understanding in the art is that there is a reasonable correlation between changes in gene expression and the level of the encoded protein. In light of the lack of support for any argument by the PTO to the contrary, Applicants submit that they have established that it is more likely than not that one of skill in the art would believe that because the PRO1557 mRNA is expressed at a higher level in esophageal tumor and kidney tumor compared to normal esophageal and normal kidney tissue, respectively, the PRO1557 polypeptide will also be expressed at a higher level in esophageal tumor and kidney tumor compared to normal esophageal and normal kidney tissue, respectively.

The Claimed Nucleic Acids would have Diagnostic Utility even if there is no Direct Correlation between Gene Expression and Protein Expression

Even assuming *arguendo* that, there is no direct correlation between changes in gene expression and changes in protein expression for PRO1557, which Applicants submit is not true, nucleic acids related to a gene that is differentially expressed in cancer would **still** have a credible, specific and substantial utility.

In paragraph 6 of the Grimaldi Declaration, Exhibit 5, Mr. Grimaldi explains that:

However, even in the rare case where the protein expression does not correlate with the mRNA expression, this still provides significant information useful for cancer diagnosis and treatment. For example, if over- or under-expression of a gene product does not correlate with over- or under-expression of mRNA in certain tumor types but does so in others, then identification of both gene expression and protein expression enables more accurate tumor classification and hence better determination of suitable therapy.

This conclusion is echoed in the Declaration of Avi Ashkenazi, Ph.D. (attached as Exhibit 12), an expert in the field of cancer biology. This declaration was previously submitted in connection with co-pending application Serial No. 09/903,925. Applicants submit that simultaneous testing of gene expression and gene product expression enables more accurate tumor classification, even if there is no positive correlation between the two. This leads to better determination of a suitable therapy.

Appl. No. : 10/063,713
Filed : May 8, 2002

This is further supported by the teachings in the article by Hanna and Mornin (attached as Exhibit 13). The article teaches that the HER-2/neu gene has been shown to be amplified and/or overexpressed in 10%-30% of invasive breast cancers and in 40-60% of intraductal breast carcinoma. Further, the article teaches that diagnosis of breast cancer includes testing both the amplification of the HER-2/neu gene (by FISH) as well as the overexpression of the HER-2/neu gene product (by IHC). Even when the protein is not overexpressed, the assay relying on both tests leads to a more accurate classification of the cancer and a more effective treatment of it.

The Applicants have established that it is the general, accepted understanding in the art that there is a positive correlation between changes in gene expression and changes in protein expression. However, even when this is not the case, a gene that is differentially expressed in cancer would still have utility. Thus, Applicants have demonstrated another basis for supporting the asserted utility for the claimed nucleic acids.

The Arguments made by the PTO are Not Sufficient to satisfy the PTO's Initial Burden of Offering Evidence "that one of ordinary skill in the art would reasonably doubt the asserted utility"

As stated above, an Applicant's assertion of utility creates a presumption of utility that will be sufficient to satisfy the utility requirement of 35 U.S.C. § 101, "unless there is a reason for one skilled in the art to question the objective truth of the statement of utility or its scope." *In re Langer*, 503 F.2d 1380, 1391, 183 USPQ 288, 297 (CCPA 1974). The evidentiary standard to be used throughout *ex parte* examination in setting forth a rejection is a preponderance of the evidence, or "more likely than not" standard. *In re Oetiker*, 977 F.2d 1443, 1445, 24 USPQ2d 1443, 1444 (Fed. Cir. 1992). This is stated explicitly in the M.P.E.P.:

[T]he applicant does not have to provide evidence sufficient to establish that an asserted utility is true "beyond a reasonable doubt." **Nor must the applicant provide evidence such that it establishes an asserted utility as a matter of statistical certainty.** Instead, evidence will be sufficient if, considered as a whole, it leads a person of ordinary skill in the art to conclude that the asserted utility is more likely than not true. M.P.E.P. at § 2107.02, part VII (2004) (underline emphasis in original, bold emphasis added, internal citations omitted).

The PTO has the initial burden to offer evidence "that one of ordinary skill in the art would reasonably doubt the asserted utility." *In re Brana*, 51 F.3d 1560, 1566, 34 U.S.P.Q.2d 1436 (Fed. Cir. 1995). Only then does the burden shift to the Applicant to provide rebuttal

Appl. No. : 10/063,713
Filed : May 8, 2002

evidence. *Id.* As stated in the M.P.E.P., such rebuttal evidence does not need to absolutely prove that the asserted utility is real. Rather, the evidence only needs to be reasonably indicative of the asserted utility.

The PTO has not offered any arguments or cited any references to establish “that one of ordinary skill in the art would reasonably doubt” that a gene differentially expressed in certain tumors can be used as a diagnostic tool. Given the lack of support for the PTO’s position, Applicants submit that the PTO has not met its initial burden of overcoming the presumption that the asserted utility is sufficient to satisfy the utility requirement. And even if the PTO has met that burden, the Applicants’ supporting rebuttal evidence is sufficient to establish that one of skill in the art would be more likely than not to believe that the claimed nucleic acids can be used as diagnostic tools for cancer, particularly esophageal and kidney cancer.

Specific Utility

The Asserted Substantial Utilities are Specific to the Claimed Nucleic Acids

Applicants next address the PTO’s assertion that the asserted utilities are not specific to the claimed nucleic acids related to PRO1557. Applicants respectfully disagree.

Specific Utility is defined as utility which is “specific to the subject matter claimed,” in contrast to “a general utility that would be applicable to the broad class of the invention.” M.P.E.P. § 2107.01 I. Applicants submit that the evidence of differential expression of the PRO1557 gene in certain types of cancer cells, along with the declarations and references discussed above, provide a specific utility for the claimed nucleic acids.

As discussed above, there are significant data which show that the gene encoding the PRO1557 polypeptide is more highly expressed in esophageal tumor tissue and kidney tumor tissue compared to normal esophageal tissue and normal kidney tissue, respectively. These data are strong evidence that the gene encoding the PRO1557 polypeptide is associated with esophageal and kidney tumors. Thus, contrary to the assertions of the PTO, Applicants submit that they have provided evidence associating the gene encoding PRO1557 with two specific diseases. The asserted utility as a diagnostic tool for cancer, particularly esophageal tumor and kidney tumor, is a specific utility – it is not a general utility that would apply to the broad class of nucleic acids.

Appl. No. : 10/063,713
Filed : May 8, 2002

Conclusion

The PTO has asserted three arguments for why there is a lack of a substantial utility: (1) the data reporting that the PRO1557 gene is differentially expressed in certain tumors is not sufficient because there is not sufficient information regarding expression levels, and because the information is too sparse to allow the encoding polynucleotide to be used; (2) that the literature cautions researchers from drawing conclusions based on small changes in transcript expression levels between normal and cancerous tissue; and, (3) that because there is no *necessary* correlation between gene amplification and protein expression, the claimed nucleic acids cannot be used as cancer diagnostic or therapeutic tools. Applicants have addressed each of these arguments in turn.

First, the Applicants provide a declaration stating that the data in Example 18 reporting higher expression of the PRO1557 gene in esophageal tumor tissue and kidney tumor tissue compared to normal esophageal tissue and normal kidney tissue, respectively, are real and significant. This declaration also indicates that given the at least two-fold difference in expression levels, the claimed nucleic acids have utility as cancer diagnostic tools. Applicants have also shown that the precise level of expression and activity or role of the PRO1557 polypeptide or the gene that encodes it in cancer is irrelevant to the utility of the claimed subject matter. Resolution of these issues is not required to use the claimed nucleic acids as tumor diagnostic tools – one does not have to know why the PRO1557 gene is differentially expressed in certain tumors to use it as a tumor marker.

Second, Applicants have shown that the Hu and Wu references cited by the PTO do not teach that genes differentially expressed in cancer cannot be used as diagnostic tools. In fact, Wu supports Applicants' asserted utility and refutes the PTO's argument based on the Hu publication.

Third, Applicants assert that whether the encoded polypeptide is also differentially expressed in certain tumors is currently not at issue in this application. However, Applicants believe that they have established that there is a reasonable correlation between changes in gene expression and corresponding changes in the level of the encoded protein. The PTO provides no evidence to the contrary. Applicants have presented the declarations of two experts in the field along with supporting references which establish that the general, accepted view of those of skill

Appl. No. : 10/063,713
Filed : May 8, 2002

in the art is that there is a direct correlation between changes in mRNA levels and the encoded protein levels.

Applicants have also presented the declarations of two experts in the field, along with supporting references, which establish that even in the anomalous case where there is no positive correlation between gene expression and expression of the encoded protein, the simultaneous monitoring of both is useful for diagnosis and further classification of the cancer.

Finally, the PTO asserts that there is no asserted specific utility. Applicants have pointed out that the substantial utilities described above are specific to the claimed nucleic acids because the gene encoding PRO1557 is differentially expressed in certain cancer cells compared to the corresponding normal cells. This is not a general utility that would apply to the broad class of nucleic acids.

Thus, given the totality of the evidence provided, Applicants submit that they have established a substantial, specific, and credible utility for the claimed nucleic acids as a diagnostic agent. According to the PTO Utility Examination Guidelines (2001), irrefutable proof of a claimed utility is not required. Rather, a specific, substantial, and credible utility requires only a “reasonable” confirmation of a real world context of use. Applicants remind the PTO that:

A small degree of utility is sufficient . . . The claimed invention must only be capable of performing some beneficial function . . . An invention does not lack utility merely because the particular embodiment disclosed in the patent lacks perfection or performs crudely . . . A commercially successful product is not required . . . Nor is it essential that the invention accomplish all its intended functions . . . or operate under all conditions . . . partial success being sufficient to demonstrate patentable utility . . . In short, **the defense of non-utility cannot be sustained without proof of total incapacity**. If an invention is only partially successful in achieving a useful result, a rejection of the claimed invention as a whole based on a lack of utility is not appropriate. M.P.E.P. at 2107.01 (underline emphasis in original, bold emphasis added, citations omitted).

Applicants submit that they have established that it is more likely than not that one of skill in the art would reasonably accept the utility for the claimed nucleic acids relating to PRO1557 set forth in the specification. In view of the above, Applicants respectfully request that the PTO reconsider and withdraw the utility rejection under 35 U.S.C. §101.

Appl. No. : 10/063,713
Filed : May 8, 2002

Rejections under 35 U.S.C. § 112, first paragraph – Enablement

The PTO rejected Claims 1-20 under 35 U.S.C. § 112, first paragraph, as containing subject matter which was not described in the specification in such a way as to enable one skilled in the art to use the invention. The PTO argues that because the claimed invention is not supported by a substantial, specific and credible utility, the claims are not enabled. The PTO states that the specification does not provide sufficient guidance or working examples to be able to use the claimed nucleic acids diagnostically or therapeutically, without undue experimentation.

Applicants respectfully traverse.

As an initial matter, Applicants submit that in the discussion of the 35 U.S.C. § 101 rejection above, Applicants have established a substantial, specific, and credible utility for the claimed nucleic acids. Applicants therefore request that the PTO reconsider and withdraw the enablement rejection to the extent that it is based on a lack of utility for the claimed nucleic acids.

As amended, the pending claims are to nucleic acids that have at least 95% or 99% nucleic acid sequence identity to the recited sequence and is “more highly expressed in esophageal tumor and kidney tumor tissue compared to normal esophageal and normal kidney tissue, respectively.” Other claimed nucleic acids can hybridize to the recited sequences under stringent conditions.

Applicants submit that the claimed nucleic acids are enabled, as one of skill in the art would know how to make and use them. It is well-established in the art how to make the claimed nucleic acids which have at least 95% or 99% sequence identity to the disclosed sequences related to SEQ ID NO: 81. Likewise, Applicants have disclosed how to determine if the recited nucleic acids are differentially expressed in esophageal and kidney tumors compared to their normal counterparts (*see, e.g.*, Example 18 beginning at paragraph [0529] of the specification). Finally, it is well-known in the art how to determine if the recited nucleic acids hybridize to the disclosed sequences under the specified stringent conditions. Thus, one of skill in the art would know how to make the claimed nucleic acids.

As discussed above, Applicants submit that they have established that one of skill in the art would believe that it is more likely than not that the PRO1557 gene is differentially expressed in esophageal and kidney tumors. Given the disclosure in the specification and the level of skill in the art, a skilled artisan would know how to use the claimed nucleic acids as diagnostic tools.

Appl. No. : 10/063,713
Filed : May 8, 2002

For example, nucleic acids which have at least 95% or 99% sequence identity to the disclosed sequences and are “more highly expressed in esophageal tumor and kidney tumor tissue compared to normal esophageal and normal kidney tissue, respectively” can be used as diagnostic tools since the claimed nucleic acids are themselves differentially expressed in certain tumors. A nucleic acid which has at least 95% or 99% sequence identity to the disclosed sequences and hybridizes to the disclosed sequences under the specified stringent conditions can be used as a hybridization probe to detect the expression of the PRO1557 gene, making it useful as a diagnostic tool. Given the skill in the art and the disclosure of how to make and use the claimed nucleic acids, Applicants request that the PTO reconsider and withdraw its rejection under 35 U.S.C. § 112, first paragraph.

Rejection under 35 U.S.C. §112, first paragraph – Written Description

The PTO has rejected Claims 1-6, 9, 10 and 14-20 under 35 U.S.C. §112, first paragraph, as containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the art that the inventors, at the time the application was filed, had possession of the invention. According to the PTO, because the claims do not require that the claimed nucleic acids or encoded polypeptides possess any particular biological activity, particular conserved structure, or other disclosed distinguishing feature, the claims fail the written description requirement. The PTO states that the claims are drawn to a genus of nucleic acids that is defined only by sequence identity. Finally, the PTO states that the only factor present in the claim is a partial structure in the form of a recitation of percent identity. The PTO concludes that in the absence of sufficient recitation of distinguishing identifying characteristics, the specification does not provide adequate written description of the claimed genus.

The Legal Standard for Written Description

The well-established test for sufficiency of support under the written description requirement of 35 U.S.C. §112, first paragraph is whether the disclosure “reasonably conveys to artisan that the inventor had possession at that time of the later claimed subject matter.” *In re Kaslow*, 707 F.2d 1366, 1375, 2121 USPQ 1089, 1096 (Fed. Cir. 1983); *see also Vas-Cath, Inc. v. Mahurkar*, 935 F.2d at 1563, 19 USPQ2d at 1116 (Fed. Cir. 1991). The adequacy of written description support is a factual issue and is to be determined on a case-by-case basis. *See e.g.*,

Appl. No. : 10/063,713
Filed : May 8, 2002

Vas-Cath, Inc. v. Mahurkar, 935 F.2d at 1563, 19 USPQ2d at 1116 (Fed. Cir. 1991). The factual determination in a written description analysis depends on the nature of the invention and the amount of knowledge imparted to those skilled in the art by the disclosure. *Union Oil v. Atlantic Richfield Co.*, 208 F.3d 989, 996 (Fed. Cir. 2000).

The Current Invention is Adequately Described

As noted above, whether the Applicants were in possession of the invention as of the effective filing date of an application is a factual determination, reached by the consideration of a number of factors, including the level of knowledge and skill in the art, and the teaching provided by the specification. The inventor is not required to describe every single detail of his/her invention. An Applicant's disclosure obligation varies according to the art to which the invention pertains. The present invention pertains to the field of recombinant DNA/protein technology. It is well-established that the level of skill in this field is very high since a representative person of skill is generally a Ph.D. scientist with several years of experience. Accordingly, the teaching imparted in the specification must be evaluated through the eyes of a highly skilled artisan as of the date the invention was made.

The subject matter of the pending claims concerns nucleic acids having 95% or 99% sequence identity to the nucleic acid sequence of SEQ ID NO:81, the full-length coding sequence of the nucleic acid sequence of SEQ ID NO:81, or the full-length coding sequence of the cDNA deposited under ATCC accession number 203317, with the functional recitation as amended: "wherein said nucleic acid is more highly expressed in esophageal tumor and kidney tumor tissue compared to normal esophageal and normal kidney tissue, respectively." Other claimed nucleic acids hybridize to the nucleic acid sequence of SEQ ID NO:81, the full-length coding sequence of the nucleic acid sequence of SEQ ID NO:81, the full-length coding sequence of the cDNA deposited under ATCC accession number 203317, or the complements thereof, under the specified stringent conditions. We turn first to the claims which recite specific high stringency hybridization conditions.

In *Enzo Biochem v. Gen-Probe Inc.*, 323 F.3d 956 (Fed. Cir. 2002), the Court held that functional descriptions of genetic material may satisfy the written description requirement. In so holding, the Court gave judicial notice to the USPTO's Manual of Patent Examining Procedure, which provides that the written description requirement may be satisfied when the disclosure

Appl. No. : 10/063,713
Filed : May 8, 2002

provides sufficiently detailed identifying characteristics, such as “complete or partial structure, other physical and/or chemical properties, *functional characteristics when coupled with a known or disclosed correlation between function and structure*, or some combination of such characteristics.” *Id.* at 964, quoting 66 Fed. Reg. at 1106 (emphasis in original). In *Enzo*, the Court found describing nucleic acids based on their ability to hybridize to another nucleic acid sequence which was adequately described may be an adequate description of the nucleic acid. This is because the hybridization function of a nucleic acid is dependent on the sequences of the nucleic acid – a disclosed function which is coupled with a known correlation between function and structure. The Court favorably discussed the PTO’s example wherein “genus claims to nucleic acids based on their hybridization properties...may be adequately described if they hybridize under highly stringent conditions to known sequences because such conditions dictate that all species within the genus will be structurally similar.” *Id.* at 967 (citing *Application of [Written Description] Guidelines*, Example 9) (emphasis added).

Applicants submit that the stringent hybridization conditions specified in the pending claims, alone or in combination with the recited percent sequence identity, result in all species within the genus being structurally similar. As the *Enzo* Court noted, Examples 9 and 10 of the Application of Written Description Guidelines (hereinafter “Guidelines”) make clear that specifying hybridization under highly stringent conditions yields “structurally similar DNAs.” Guidelines, Example 9 at page 36. The analysis of a genus claim in Example 10 of the Guidelines states:

[T]urning to the genus analysis, the art indicates that *there is no substantial variation within the [claimed] genus because of the stringency of hybridization conditions which yields structurally similar molecules*. The single disclosed species is representative of the genus because reduction to practice of this species, considered along with the defined hybridization conditions and the level of skill and knowledge in the art, are sufficient to allow the skilled artisan to recognize that applicant was in possession of the necessary common attributes or features of the elements possessed by the members of the genus. Guidelines, Example 10 at page 39 (emphasis added).

Given the level of skill in the art, specifying highly stringent conditions leads to “no substantial variation within the [claimed] genus,” and therefore a skilled artisan would recognize that the Applicants were in possession of the necessary common attributes or features of the genus. This is contrary to the PTO’s argument the claimed sequences do not possess “any

Appl. No. : 10/063,713
Filed : May 8, 2002

particular conserved structure, or other disclosed distinguishing feature.” Office Action at 11. The common element or attribute of the claimed genus of nucleic acids is that species of the genus contain a nucleic acid which is structurally related to SEQ ID NO: 81, such that the nucleic acids hybridize to SEQ ID NO: 81 or the related sequences under the specified high stringency conditions recited in the claims.

The present situation is not analogous to *Fiddes v. Baird*, 30 U.S.P.Q. 2d 1481, cited by the PTO. Unlike *Fiddes*, where arguably the structure of other mammalian sequences could not be conceived based on a single species of the genus, here the skill in the art is such that the sequence of nucleic acids which hybridize to SEQ ID NO: 81 under the conditions specified can be conceived. Here, the claimed genus is defined by its structure – members of the genus hybridize under the specified conditions to the specified sequences, each of which are adequately described in the specification.

Applicants submit that the pending claims relating to nucleic acids having 95% or 99% sequence identity to the nucleic acids related to SEQ ID NO:81 with the functional recitation “wherein said nucleic acid is more highly expressed in esophageal tumor and kidney tumor tissue compared to normal esophageal and normal kidney tissue, respectively” are also adequately described. In Example 14 of the written description training materials, the written description requirement was found to be satisfied for claims relating to polypeptides having 95% homology to a particular sequence and possessing a particular catalytic activity, even though the applicant had not made any variants. Similarly, the pending claims also have very high sequence homology to the disclosed sequences and must share the same expression pattern in certain tumors. In Example 14, the procedures for making variants were known in the art and the disclosure taught how to test for the claimed catalytic activity. Similarly, in the instant application, it is well known in the art how to make nucleic acids which have at least 95% sequence identity to the disclosed sequences, and the specification discloses how to test to determine if the nucleic acid sequence is differentially expressed in esophageal or kidney tumors. Like Example 14, the genus of nucleic acids that have at least 95% or 99% sequence identity to the disclosed sequences will not have substantial variation since all of the variants must have the same expression in certain tumors.

Furthermore, while Applicants appreciate that actions taken by the PTO in other applications are not binding with respect to the examination of the present application,

Appl. No. : 10/063,713
Filed : May 8, 2002

Applicants note that the PTO has issued many patents containing claims to variant nucleic acids or variant proteins where the applicants did not actually make such nucleic acids or proteins. Representative patents include U.S. Patent No. 6,737,522, U.S. Patent No. 6,395,306, U.S. Patent No. 6,025,156, U.S. Patent No. 6,645,499, U.S. Patent No. 6,498,235, and U.S. Patent No. 6,730,502, which are submitted herewith as Exhibits 14-19.

In conclusion, Applicants submit that they have satisfied the written description requirement for the pending claims based on the actual reduction to practice of SEQ ID NO: 81, by specifying the high stringency conditions under which hybridization occurs, and by describing the gene expression assay, all of which result in a lack of substantial variability in the species falling within the scope of the instant claims. Applicants submit that this disclosure would allow one of skill in the art to "recognize that the applicant was in possession of the necessary common attributes or features of the elements possessed by the members of the genus." Hence, Applicants respectfully request that the PTO reconsider and withdraw the written description rejection under 35 U.S.C. §112.

Rejections under 35 U.S.C. § 112, second paragraph – Indefiniteness

The PTO has rejected Claim 15 under 35 U.S.C. § 112, second paragraph, as being indefinite. The PTO objects to the use of "stringent conditions." Claim 15 has been canceled.

The PTO has also rejected Claims 1-6, 9, 10, and 14, and dependent claims 7, 8, 11-13, 16 and 17-20, under 35 U.S.C. § 112, second paragraph, as being indefinite. The PTO objects to the recitation of "the extracellular domain" allegedly because no extracellular domain is identified.

Applicants have amended the claims to delete any reference to an extracellular domain. Claim 14 is further rejected as indefinite because the conditions for hybridization of the claimed nucleic acid are allegedly unclear. Claim 14 is amended to recite stringent conditions for hybridization of the claimed nucleic acid. In light of these amendments, Applicants request that the PTO withdraw the indefiniteness rejections under 35 U.S.C. §112, second paragraph.

Appl. No. : 10/063,713
Filed : May 8, 2002

Rejection under 35 U.S.C. §102(b) – Anticipation

The PTO has rejected Claims 14-16 as anticipated under 35 U.S.C. §102(b) by GenBank Accession AA040433. According to the PTO, AA040433 discloses a nucleic acid that is 97% identical to SEQ ID NO:81 over 381 consecutive bases.

While Applicants do not acquiesce to the PTO's position that the nucleotide sequence of AA040433 is encompassed by the polynucleotide of Claim 14, Applicants have canceled Claims 15, and amended Claims 14 and 16 such that the recited nucleic acid must be at least 450 nucleotides, or at least about 500 nucleotides in length, respectively. The polynucleotide disclosed in AA040433 does not anticipate amended Claims 14 or 16. Accordingly, Applicants request that the PTO reconsider and withdraw the rejection of Claims 14 and 16 under 35 U.S.C. § 102(b).

The PTO has rejected Claims 1-10 and 12-20 as anticipated under 35 U.S.C. §102(b) by WO 00/70049. The PTO states that WO 00/70049 teaches a nucleotide sequence that is 1720 nucleotide in length and 100% identical to nucleotides 1-1720 of Applicants' SEQ ID NO:81. As discussed above, the instant claimed subject matter has utility based upon the data in Example 18 and the instant application is a continuation of PCT/US00/23328; therefore, the present claims are entitled to the filing date of August 24, 2000. WO 00/70049 is not prior art under § 102(b).

WO 00/70049 was published on November 23, 2000, which is subsequent to the filing of priority application PCT/US00/23328 (August 24, 2000). Again, PCT/US00/23328 discloses the differential expression data which provides utility for the instant claims, and Applicants are entitled to the filing date of August 24, 2000. Therefore, WO 00/70049 cannot be cited under § 102(b).

In view of the above discussion, reconsideration and withdrawal of the rejection under § 102(b) is respectfully requested

Appl. No. : 10/063,713
Filed : May 8, 2002

CONCLUSION

In view of the above, Applicants respectfully maintain that claims are patentable and request that they be passed to issue. Applicants invite the Examiner to call the undersigned if any remaining issues may be resolved by telephone.

Please charge any additional fees, including any fees for additional extension of time, or credit overpayment to Deposit Account No. 11-1410.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: April 12, 2005

By: 

AnneMarie Kaiser
Registration No. 37,649
Attorney of Record
Customer No. 30,313
(619) 235-8550

1520828_1
040205

Appl. No. : **10/063,713**
Filed : **May 8, 2002**

DELETION OF INVENTORS

Please correct the inventorship under 37 CFR §1.48(b) by removing the following inventors from the present application:

Dan L. Eaton, Ellen Filvaroff, Mary E. Gerritsen, and Colin K. Watanabe.

Applicants request that these inventors be deleted, as their inventions are no longer being claimed in the present application as a result of prosecution.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	: Eaton, et al.
Appl. No.	: 10/063,557
Filed	: May 2, 2002
For	: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	: David J. Blanchard
Group Art Unit	: 1642

DECLARATION OF J. CHRISTOPHER GRIMALDI, UNDER 37 CFR §1.132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

I, J. Christopher Grimaldi, declare and state as follows:

1. I am a Senior Research Associate in the Molecular Biology Department of Genentech, Inc., South San Francisco, CA 94080.
2. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).
3. I joined Genentech in January of 1999. From 1999 to 2003, I directed the Cloning Laboratory in the Molecular Biology Department. During this time I directed or performed numerous molecular biology techniques including semi-quantitative Polymerase Chain Reaction (PCR) analyses. I am currently involved, among other projects, in the isolation of genes coding for membrane associated proteins which can be used as targets for antibody therapeutics against cancer. In connection with the above-identified patent application, I personally performed or directed the semi-quantitative PCR gene expression analyses in the assay entitled "Tumor Versus Normal Differential Tissue Expression Distribution," which is described in EXAMPLE 18 in the specification. These studies were used to identify differences in gene expression between tumor tissue and their normal counterparts.
4. EXAMPLE 18 reports the results of the PCR analyses conducted as part of the investigating of several newly discovered DNA sequences. This process included developing

Appl. No. : 10/063,557
Filed : May 2, 2002

primers and analyzing expression of the DNA sequences of interest in normal and tumor tissues. The analyses were designed to determine whether a difference exists between gene expression in normal tissues as compared to tumor in the same tissue type.

5. The DNA libraries used in the gene expression studies were made from pooled samples of normal and of tumor tissues. Data from pooled samples is more likely to be accurate than data obtained from a sample from a single individual. That is, the detection of variations in gene expression is likely to represent a more generally relevant condition when pooled samples from normal tissues are compared with pooled samples from tumors in the same tissue type.

6. In differential gene expression studies, one looks for genes whose expression levels differ significantly under different conditions, for example, in normal versus diseased tissue. Thus, I conducted a semi-quantitative analysis of the expression of the DNA sequences of interest in normal versus tumor tissues. Expression levels were graded according to a scale of +, -, and +/- to indicate the amount of the specific signal detected. Using the widely accepted technique of PCR, it was determined whether the polynucleotides tested were more highly expressed, less expressed, or whether expression remained the same in tumor tissue as compared to its normal counterpart. Because this technique relies on the visual detection of ethidium bromide staining of PCR products on agarose gels, it is reasonable to assume that any detectable differences seen between two samples will represent at least a two fold difference in cDNA.

7. The results of the gene expression studies indicate that the genes of interest can be used to differentiate tumor from normal. The precise levels of gene expression are irrelevant; what matters is that there is a relative difference in expression between normal tissue and tumor tissue. The precise type of tumor is also irrelevant; again, the assay was designed to indicate whether a difference exists between normal tissue and tumor tissue of the same type. If a difference is detected, this indicates that the gene and its corresponding polypeptide and antibodies against the polypeptide are useful for diagnostic purposes, to screen samples to differentiate between normal and tumor. Additional studies can then be conducted if further information is desired.

8. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By:


J. Christopher Grimaldi

Date:

8/10/2004

EXHIBIT A

J. Christopher Grimaldi

1434-36th Ave.
San Francisco, CA 94122
(415) 681-1639 (Home)

EDUCATION

University of California, Berkeley
Bachelor of Arts in Molecular Biology, 1984

EMPLOYMENT EXPERIENCE

SRA

Genentech Inc., South San Francisco; 1/99 to present

Previously, was responsible to direct and manage the Cloning Lab. Currently focused on isolating cancer specific genes for the Tumor Antigen (TAP), and Secreted Tumor Protein (STOP) projects for the Oncology Department as well as Immunologically relevant genes for the Immunology Department. Directed a lab of 6 scientists focused on a company-wide team effort to identify and isolate secreted proteins for potential therapeutic use (SPDI). For the SPDI project my duties were, among other things, the critically important coordination of the cloning of thousands of putative genes, by developing a smooth process of communication between the Bioinformatics, Cloning, Sequencing, and Legal teams. Collaborated with several groups to discover novel genes through the Curagen project, a unique differential display methodology. Interacted extensively with the Legal team providing essential data needed for filing patents on novel genes discovered through the SPDI, TAP and Curagen projects. My group has developed, implemented and patented high throughput cloning methodologies that have proven to be essential for the isolation of hundreds of novel genes for the SPDI, TAP and Curagen projects as well as dozens of other smaller projects.

Scientist

DNAX Research Institute, Palo Alto; 9/91 to 1/99

Involved in multiple projects aimed at understanding novel genes discovered through bioinformatics studies and functional assays. Developed and patented a method for the specific depletion of eosinophils in vivo using monoclonal antibodies. Developed and implemented essential technical methodologies and provided strategic direction in the areas of expression, cloning, protein purification, general molecular biology, and monoclonal antibody production. Trained and supervised numerous technical staff.

Facilities Manager

Corixa, Redwood City; 5/89 - 7/91.

Directed plant-related activities, which included expansion planning, maintenance, safety, purchasing, inventory control, shipping and receiving, and laboratory management. Designed and implemented the safety program. Also served as liaison to regulatory agencies at the local, state and federal level. Was in charge of property leases, leasehold improvements, etc. Negotiated vendor contracts and directed the purchasing department. Trained and supervised personnel to carry out the above-mentioned duties.

SRA University of California, San Francisco
Cancer Research Institute; 2/87-4/89.

Was responsible for numerous cloning projects including: studies of somatic hypermutation, studies of AIDS-associated lymphomas, and cloning of t(5;14), t(11;14), and t(8;14) translocations. Focused on the activation of hemopoietic growth factors involved in the t(5;14) translocation in leukemia patients..

Research Berlex Biosciences, South San Francisco; 7/85-2/87.
Technician

Worked on a subunit porcine vaccine directed against *Mycoplasma hyopneumoniae*. Was responsible for generating genomic libraries, screening with degenerate oligonucleotides, and characterizing and expressing clones in *E. coli*. Also constructed a general purpose expression vector for use by other scientific teams.

PUBLICATIONS

1. Hilary F. Clark, et al. "The Secreted Protein Discovery Initiative (SPDI), a Large-scale Effort to Identify Novel Human Secreted and Transmembrane Proteins: a bioinformatics assessment." *Genome Res.* Vol 13(10), 2265-2270, 2003
2. Sean H. Adams, Clarissa Chui, Sarah L. Schilbach, Xing Xian Yu, Audrey D. Goddard, J. Christopher Grimaldi, James Lee, Patrick Dowd, David A. Lewin, & Steven Colman "BFTT, a Unique Acyl-CoA Thioesterase Induced in Thermogenic Brown Adipose Tissue: Cloning, organization of the human gene and assessment of a potential link to obesity" *Biochemical Journal*, Vol 360, 135-142, 2001
3. Szeto W, Jiang W, Tice DA, Rubinfeld B, Hollingshead PG, Fong SE, Dugger DL, Pham T, Yansura D, Wong TA, Grimaldi JC, Corpuz RT, Singh JS, Frantz GD, , Devaux B, Crowley CW, Schwall RH, Eberhard DA, Rastelli L, Polakis P, and Pennica D. "Overexpression of the Retenoic Acid-Responsive Gene *Stra6* in Human Cancers and its Synergistic Activation by Wnt-1 and Retinoic Acid." *Cancer Research* Vol. 61(10), 4197-4205, 2001
4. Jeanne Kahn, Fuad Mehraban, Gladdys Ingle, Xiaohua Xin, Juliet E. Bryant, Gordon Vehar, Jill Schoenfeld, J. Christopher Grimaldi (incorrectly named as "Grimaldi, CJ"), Franklin Peale, Aparna Draksharapu, David A. Lewin, and Mary E. Gerritsen. "Gene Expression Profiling in an in Vitro Model of Angiogenesis." *American Journal of Pathology* Vol 156(6), 1887-1900, 2000.
5. Grimaldi JC, Yu NX, Grunig G, Seymour BW, Cottrez F, Robinson DS, Hosken N, Ferlin WG, Wu X, Soto H, O'Garra A, Howard MC, Coffman RL. "Depletion of eosinophils in mice through the use of antibodies specific for C-C chemokine receptor 3 (CCR3). *Journal of Leukocyte Biology*; Vol. 65(6), 846-53, 1999
6. Oliver AM, Grimaldi JC, Howard MC, Kearney JF. "Independently ligating CD38 and Fc gammaRIIB relays a dominant negative signal to B cells." *Hybridoma* Vol. 18(2), 113-9, 1999

7. Cockayne DA, Muchamuel T, Grimaldi JC, Muller-Steffner H, Randall TD, Lund FE, Murray R, Schuber F, Howard MC. "Mice deficient for the ecto-nicotinamide adenine dinucleotide glycohydrolase CD38 exhibit altered humoral immune responses." *Blood* Vol. 92(4), 1324-33, 1998
8. Frances E. Lund, Nanette W. Solvason, Michael P. Cooke, Andrew W. Heath, J. Christopher Grimaldi, Troy D. Randall, R. M. E. Parkhouse, Christopher C Goodnow and Maureen C. Howard. "Signaling through murine CD38 is impaired in antigen receptor unresponsive B cells." *European Journal of Immunology*, Vol. 25(5), 1338-1345, 1995
9. M. J. Guimaraes, J. F. Bazan, A. Zolotnik, M. V. Wiles, J. C. Grimaldi, F. Lee, T. McClanahan. "A new approach to the study of haematopoietic development in the yolk sac and embryoid body." *Development*, Vol. 121(10), 3335-3346, 1995
10. J. Christopher Grimaldi, Sriram Balasubramanian, J. Fernando Bazan, Armen Shanafelt, Gerard Zurawski and Maureen Howard. "CD38-mediated protein ribosylation." *Journal of Immunology*, Vol. 155(2), 811-817, 1995
11. Leopoldo Santos-Argumedo, Frances F. Lund, Andrew W. Heath, Nanette Solvason, Wei Wei Wu, J. Christopher Grimaldi, R. M. E. Parkhouse and Maureen Howard. "CD38 unresponsiveness of xid B cells implicates Bruton's tyrosine kinase (btk) as a regulator of CD38 induced signal transduction." *International Immunology*, Vol 7(2), 163-170, 1995
12. Frances Lund, Nanette Solvason, J. Christopher Grimaldi, R. M. E. Parkhouse and Maureen Howard. "Murine CD38: An immunoregulatory ectoenzyme." *Immunology Today*, Vol. 16(10), 469-473, 1995
13. Maureen Howard, J. Christopher Grimaldi, J. Fernando Bazan, Frances E. Lund, Leopoldo Santos-Argumedo, R. M. E. Parkhouse, Timothy F. Walseth, and Hon Cheung Lee. "Formation and Hydrolysis of Cyclic ADP-Ribose Catalyzed by Lymphocyte Antigen CD38." *Science*, Vol. 262, 1056-1059, 1993
14. Nobuyuki Harada, Leopoldo Santos-Argumedo, Ray Chang, J. Christopher Grimaldi, Frances Lund, Camilynn I. Brannan, Neal G. Copeland, Nancy A. Jenkins, Andrew Heath, R. M. E. Parkhouse and Maureen Howard. "Expression Cloning of a cDNA Encoding a Novel Murine B Cell Activation Marker: Homology to Human CD38." *The Journal of Immunology*, Vol. 151, 3111-3118, 1993
15. David J. Rawlings, Douglas C. Saffran, Satoshi Tsukada, David A. Largaespada, J. Christopher Grimaldi, Lucie Cohen Randolph N. Mohr, J. Fernando Bazan, Maureen Howard, Neal G. Copeland, Nancy A. Jenkins, Owen Witte. "Mutation of Unique Region of Bruton's Tyrosine Kinase in Immunodeficient XID Mice." *Science*, Vol. 261, 358-360, 1993
16. J. Christopher Grimaldi, Raul Torres, Christine A. Kozak, Ray Chang, Edward Clark, Maureen Howard, and Debra A. Cockayne. "Genomic Structure and Chromosomal Mapping of the Murine CD40 Gene." *The Journal of Immunology*, Vol 149, 3921-3926, 1992
17. Timothy C. Meeker, Bruce Shiramizu, Lawrence Kaplan, Brian Herndier, Henry Sanchez, J. Christopher Grimaldi, James Baumgartner, Jacob Rachlin, Ellen Feigal, Mark Rosenblum and Michael S. McGrath. "Evidence for Molecular Subtypes of HIV-Associated Lymphoma:

Division into Peripheral Monoclonal, Polyclonal and Central Nervous System Lymphoma." AIDS, Vol. 5, 669-674, 1991

18. Ann Grimaldi and Chris Grimaldi. "Small-Scale Lambda DNA Prep." Contribution to Current Protocols in Molecular Biology, Supplement 5, Winter 1989
19. J. Christopher Grimaldi, Timothy C. Meeker. "The t(5;14) Chromosomal Translocation in a Case of Acute Lymphocytic Leukemia Joins the Interleukin-3 Gene to the Immunoglobulin Heavy Chain Gene." Blood, Vol. 73, 2081-2085, 1989
20. Timothy C. Meeker, J. Christopher Grimaldi, et al. "An Additional Breakpoint Region in the BCL-1 Locus Associated with the t(11;14) (q13;q32) Translocation of B-Lymphocytic Malignancy." Blood, Vol. 74, 1801-1806, 1989
21. Timothy C. Meeker, J. Christopher Grimaldi, Robert O'Rourke, et al. "Lack of Detectable Somatic Hypermutation in the V Region of the Ig H Chain Gene of a Human Chronic B Lymphocytic Leukemia." The Journal of Immunology, Vol. 141, 3994-3998, 1988

MANUSCRIPTS IN PREPARATION

1. Sriram Balasubramanian, J. Christopher Grimaldi, J. Fernando Bazan, Gerard Zurawski and Maureen Howard. "Structural and functional characterization of CD38: Identification of active site residues"

PATENTS

1. "Methods for Eosinophil Depletion with Antibody to CCR3 Receptor" (US 6,207,155 B1).
2. "Amplification Based Cloning Method." (US 6,607,899)
3. Ashkenazi et al., "Secreted and Transmembrane Polypeptides and Nucleic Acids Encoding the Same." (this patent covers several hundred genes)
4. "IL-17 Homologous Polypeptides and Therapeutic Uses Thereof"
5. "Method of Diagnosing and Treating Cartilaginous Disorders."

MEMBERSHIPS AND ACTIVITIES

Editor	Frontiers in Bioscience
Member	DNAX Safety Committee 1991-1999
	Biological Safety Affairs Forum (BSAF) 1990-1991
	Environmental Law Foundation (ELF) 1990-1991



US006465185B1

(12) **United States Patent**
Goldfine et al.

(10) Patent No.: **US 6,465,185 B1**
(45) Date of Patent: **Oct. 15, 2002**

(54) **POLYMORPHIC HUMAN PC-1 SEQUENCES ASSOCIATED WITH INSULIN RESISTANCE**

(75) Inventors: **Ira Goldfine**, San Francisco, CA (US); **Vincenzo Trischitta**, San Giovanni Rotondo (IT); **Riccardo Vigneri**, Catania (IT); **Antonio Pizzuti**, San Giovanni Rotondo (IT); **Lucia A. Frittitta**, Catania (IT)

(73) Assignees: **Istituto di Ricovero e Cura a Carattere Scientifico**, Foggia (IT); **The Regents of the University of California**, Oakland, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/438,906**

(22) Filed: **Nov. 18, 1999**

Related U.S. Application Data

(60) Provisional application No. 60/108,853, filed on Nov. 18, 1998.

(51) Int. Cl.⁷ **C12Q 1/68; C07H 21/04**

(52) U.S. Cl. **435/6; 536/23.1; 536/24.3; 536/24.33; 536/24.31**

(58) Field of Search **435/6; 536/23.1; 536/24.31, 24.33, 24.3**

(56) **References Cited**

PUBLICATIONS

Buckley et al. Genbank Accession No. M57736, Mar. 1995.*
Hillier et al. Genbank Accession No. N94942, Apr. 1996.*
Trischitta et al. "Early molecular defects in human insulin resistance" *Diabetes Met.* vol. 13, No. 3, p. 147-162, 1997.*
Whitehead et al. "Expression of the putative inhibitor of the insulin receptor tyrosine kinase PC-1 in dermal fibroblasts from patients with syndromes of severe insulin resistance" *Clin. Endocrin.* vol. 47, p. 65-70, Jul. 1997.*
Rasmussen et al. "The K121Q variant of the human PC-1 gene is not associated with insulin resistance" *Diabetes*, vol. 49, p. 1608-1611, Sep. 2000.*
Accession No. AA621565.

Accession No. AI090484.

Accession No. D12485 and DO1250.

Accession No. M57736 and J05654.

Argolias, A., et al., "Role of PC-1 Gene Polymorphism (K121Q) in the Progression of Renal Disease in Type 1 Diabetic Patients," *Diabetologia* (Aug. 1999) vol. 42:A267.

Belli, et al., "Identification and Characterization of a Soluble Form of the Plasma Cell Membrane Glycoprotein PC-1 (5'-nucleotide phosphodiesterase)," *J. Biochem* (1993) vol.217:421-428.

Frittitta, L., et al., "K121Q Polymorphism in Exon 4 of PC-1 Gene is Strongly Associated With Insulin Resistance," *Diabetologia* (Aug. 1999) vol. 42:A20.

Frittitta, Lucia, et al., "Elevated PC-1 Content in Cultured Skin Fibroblasts Correlates With Decreased In Vivo and In Vitro Insulin Action in Nondiabetic Subjects," *Diabetes* (Jul. 1998) vol. 47:1095-1100.

Goldfine, Ira D., et al., "Membrane Glycoprotein PC-1 and Insulin Resistance," *Molecular and Cellular Biochemistry* (1988) vol. 182:177-184.

Pizzuti, Antonio, et al., "A Polymorphism (K121Q) of the Human Glycoprotein PC-1 Gene Coding Is Strongly Associated With Insulin Resistance," *Diabetes* (Sep. 1999) vol. 48:1881-1884.

Reaven, Gerald M., "Role of Insulin Resistance in Human Disease," *Diabetes* (Dec. 1988) vol. 37:1595-1607.

* cited by examiner

Primary Examiner—W. Gary Jones

Assistant Examiner—Jeanine Goldberg

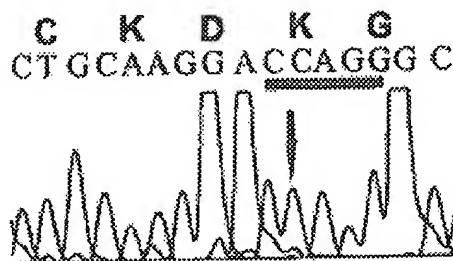
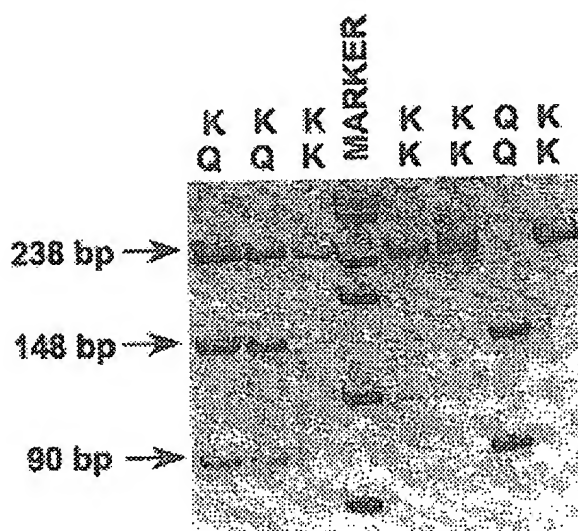
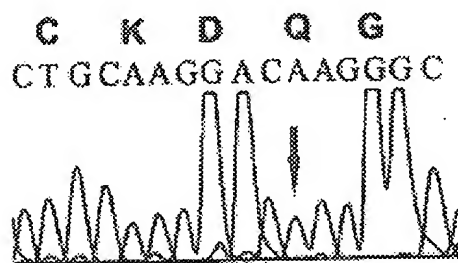
(74) *Attorney, Agent, or Firm*—Pamela J. Sherwood; James S. Keddie; Bozicevic, Field & Francis LLP

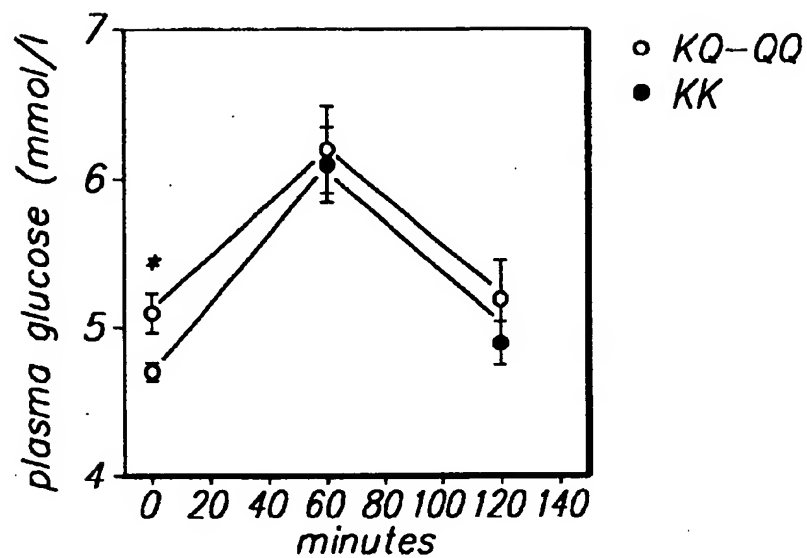
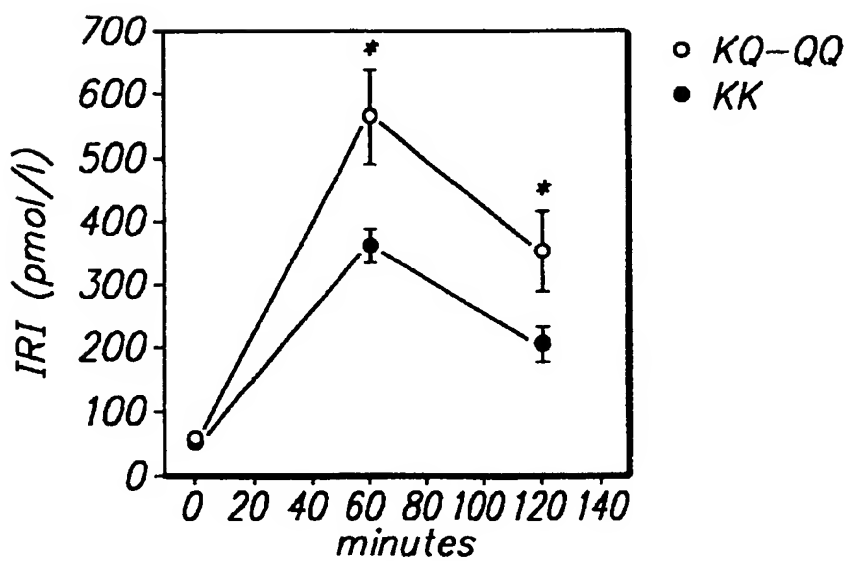
(57)

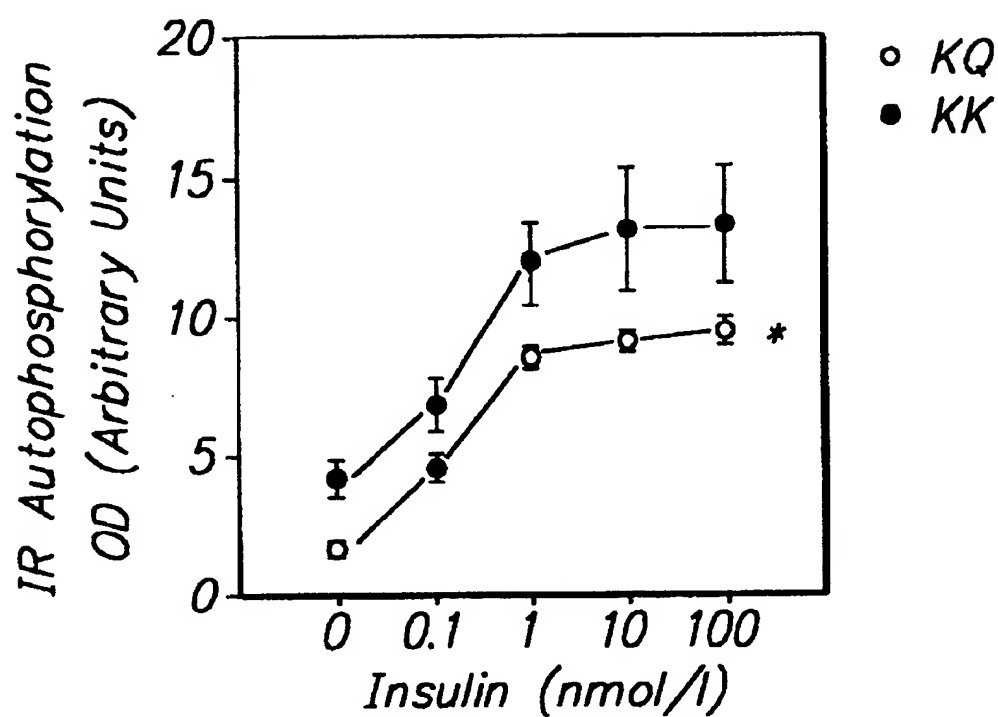
ABSTRACT

A novel polymorphism in the human PC-1 gene is characterized, which is associated with an increased predisposition to developing insulin resistance. The polymorphism affects heterozygous and homozygous carriers of the allele. The subject nucleic acids and fragments thereof, encoded polypeptides, and antibodies specific for the polymorphic amino acid sequence are useful in determining a genetic predisposition to insulin resistance. The encoded protein is useful in drug screening for compositions that affect the activity of PC-1 and insulin receptor activity or expression.

3 Claims, 3 Drawing Sheets

**FIG. 1A****FIG. 1B**

**FIG. 2A****FIG. 2B**

**FIG. 3**

POLYMORPHIC HUMAN PC-1 SEQUENCES ASSOCIATED WITH INSULIN RESISTANCE

CROSS-REFERENCE

This application claims priority to provisional patent application No. 60/108,853, filed Nov. 18, 1998.

Insulin resistance occurs in 25% of non-diabetic, non-obese, apparently healthy individuals, and predisposes them to both diabetes and coronary artery disease. Diabetes mellitus is a major health problem in the United States affecting approximately 7% of the population. The most common form of diabetes mellitus is non-insulin-dependent diabetes mellitus (NIDDM or type II diabetes). Hyperglycemia in type II diabetes is the result of both resistance to insulin in muscle and other key insulin target tissues, and decreased beta cell insulin secretion. Longitudinal studies of individuals with a strong family history of diabetes indicate that the insulin resistance precedes the secretory abnormalities. Prior to developing diabetes these individuals compensate for their insulin resistance by secreting extra insulin. Diabetes results when the compensatory hyperinsulinemia fails. The secretory deficiency of pancreatic beta cells then plays a major role in the severity of the diabetes.

Reaven (1988) *Diabetes* 37:1595-607 were the first to have investigated insulin resistant, non-diabetic, healthy individuals from the general population who are non-obese. Strikingly, they observed that 25% of them have insulin resistance that is of a similar magnitude to that seen in type II diabetes patients. These individuals compensate by having insulin levels that are 3-4 times higher than normal. These elevated insulin levels are sufficient to maintain normoglycemia. Others have also confirmed that a large proportion of the non-diabetic population is insulin resistant. These insulin resistant, non-diabetic individuals have a much higher risk for developing type II diabetes than insulin sensitive subjects.

However, even without developing hyperglycemia and diabetes, these insulin resistant individuals pay a significant price in terms of general health. Insulin resistance results in an increased risk for having elevated plasma triglycerides (TG), lower high density lipoproteins (HDL), and high blood pressure, a cluster of abnormalities that have been termed by different investigators as either Syndrome X, the insulin resistance syndrome, or the metabolic syndrome. It is believed that either the hyperinsulinemia, insulin resistance, or both play a direct role in causing these abnormalities. Data from ethnic, family, and longitudinal studies suggest that a major component of resistance is inherited.

The cellular response to insulin is mediated through the insulin receptor (IR), which is a tetrameric protein consisting of two identical extracellular alpha subunits that bind the hormone and two identical transmembrane beta subunits that have intracellular tyrosine kinase activity. When insulin binds to the IR alpha subunit, the beta subunit tyrosine kinase domain is activated, and insulin action ensues. When insulin activates the receptor, the beta subunit is autophosphorylated at the juxtamembrane domain, the tyrosine kinase domain and the C-terminal domain. Subsequently, endogenous substrates including IRS-1, IRS-2 and SHC are tyrosine phosphorylated. These phosphorylated substrates act as docking molecules to activate SH2 domain molecules including: GRB-2 which activates the ras pathway; the p85 subunit of PI-3-kinase; protein tyrosine phosphatase PTP2/SYP; PLC γ /NCK; AKT and others.

PC-1 is a class II transmembrane glycoprotein that is located both on plasma membranes and in the endoplasmic

reticulum. PC-1 is the same protein as liver nucleotide pyrophosphatase/alkaline phosphodiesterase I. In addition to muscle tissue, PC-1 has been reported to be expressed in plasma and intracellular membranes of plasma cells, placenta, the distal convoluted tubule of the kidney, ducts of the salivary gland, epididymis, proximal part of the vas deferens, chondrocytes and dermal fibroblasts. PC-1 exists as a disulfide linked homodimer of 230-260 kDa; the reduced form of the protein has a molecular size of 115-135 kDa, depending on the cell type. Human PC-1 is predicted to have 873 amino acids.

PC-1 is inserted into the plasma membrane such that there is a small cytoplasmic amino terminus, and a larger extracellular carboxyl terminus. The extracellular domain of PC-1 has a high cysteine region that is involved in dimer formation, an ATP binding site and enzymatic activity which cleaves sugar-phosphate, phosphosulfate, pyrophosphate, and phosphodiesterase linkages. The active enzyme site for phosphodiesterase and pyrophosphatase contains a key threonine residue, however a mutation of this residue does not impair the ability of PC-1 to inhibit IR function.

Belli et al. (1993) *Eur J Biochem* 217(1):421-8 discloses the existence of enzymatically active water-soluble forms of PC-1. Biosynthetic studies revealed a single, monomeric, endoglycosidase-H-sensitive membrane PC-1 precursor, which was gradually converted to a disulphide-bonded, endoglycosidase-H-resistant form. The soluble form of PC-1 does not appear to arise by proteolytic cleavage from the cell surface, although cleavage inside the cell remains a possibility. The data suggest that the most likely site of cleavage is between Pro 152 and Ala 153.

PC-1 levels are increased in fibroblasts from most patients with typical NIDDM and insulin resistance. In addition, overexpression of PC-1 in transfected cultured cells reduces insulin-stimulated tyrosine kinase activity (Goldfine et al. (1998) *Mol Cell Biochem* 182:177-184). PC-1 content in fibroblasts negatively correlates with both decreased in vivo insulin sensitivity and decreased in vitro IR autophosphorylation (Frittitta et al. (1998) *Diabetes* 47:1095-1100).

In cells from insulin-resistant subjects, insulin stimulation of glycogen synthetase was decreased. PC-1 content is also elevated in fibroblasts, muscle and fat of non-diabetic insulin resistant subjects. The elevation of PC-1 content may be a primary factor in the cause of insulin resistance, although the mechanism by which PC-1 inhibits insulin receptor activity is unknown.

Many mechanisms may potentially contribute to insulin resistance. One major mechanism is the impairment of insulin receptor tyrosine kinase (IR-TK) activity, a key step in insulin receptor signalling. Several inhibitors of IR-TK have been associated to insulin resistance. Among them is PC-1, a class II transmembrane glycoprotein that is overexpressed in tissues of insulin resistant subjects. The human PC-1 gene has been assigned to the same chromosomal region (6q22-q23) where both STS D6S290 (which has been linked to type 2 diabetes in Mexican-Americans), and the gene responsible for transient neonatal diabetes map. The identification and characterization of genetic sequences involved in insulin resistance is of great medical interest.

Database References for Genetic Sequences

The human cDNA and encoded amino acid sequence for PC-1 may be accessed in Genbank, M57736 J05654. As a reference, the "K" allele is provided herein as SEQ ID NO:1, and the encoded polypeptide as SEQ ID NO:2. The "Q" allele is provided as SEQ ID NO:3, and the encoded polypeptide as SEQ ID NO:4.

SUMMARY OF THE INVENTION

Human PC-1 nucleic acids and polypeptides are provided, including promoter and intron-exon boundaries. Polymorphic sequences are provided that encode a form of the protein associated with increased insulin resistance, where a naturally occurring polymorphism of interest comprises a lys→glu substitution at position 121 of the protein, in the high cysteine region. Also provided are polymorphisms in the 3' untranslated region of PC-1. The subject nucleic acids and fragments thereof, encoded polypeptides, and antibodies specific for the polymorphic amino acid sequence are useful in determining a genetic predisposition to insulin resistance. The encoded protein is useful in drug screening for compositions that affect the activity of PC-1 and insulin receptor activity or expression. Screening methods that analyze plasma levels of soluble PC-1 are also provided, where convenient quantitation of PC-1 content is used in diagnosis of insulin resistance.

BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1A and 1B. Sequence analysis of PC-1 exon 4. Arrows point to nucleotide N. The Q allele sequence is depicted in the upper panel, the K allele in the lower. The Avail restriction enzyme recognition site is underlined in the Q allele sequence. Avail digestion of PC-1 exon 4 amplimers from 7 different genomic DNAs. The 238 bp amplimer is completely digested in the QQ sample, resulting two smaller 148 and 90 bp fragments. While KK samples remain undigested, KQ samples reveal a partial (50%) digestion.

FIGS. 2(A) and 2(B). Plasma glucose (A) and insulin (B) levels during an OGTT (75 g) in Q allele carriers (n=33, white circles) and KK subjects (n=68, black circles). §=p<0.05 and *=p<0.01 vs. KK subjects.

FIG. 3. Insulin receptor autophosphorylation in fibroblasts from Q allele carriers (n=5, white circles) and KK subjects (n=5, black circles). This function was determined by exposing cells for 10 minutes to increasing insulin concentrations (0–100 nM). Cells were then solubilized and the insulin receptor immunocaptured on plastic wells pre-coated with a monoclonal antibody specific to the insulin receptor. After washing, a biotinylated antiphosphotyrosine antibody was added followed by peroxidase-conjugated streptavidin detection assay. Data are expressed as arbitrary units, normalized for protein content. **=p<0.01 vs. KK subjects.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

Methods and compositions are provided for diagnosing a predisposition to human insulin resistance. The methods comprise an analysis of germline DNA for a predisposing polymorphism in the gene encoding PC-1, where presence of the altered gene confers an increased susceptibility to insulin resistance. Human PC-1 gene and gene product compositions are provided that encode specific polymorphic forms of PC-1. Polymorphisms of interest include a coding change at position 121 of the protein, and polymorphisms of the 3' UTR.

In another embodiment of the invention, the concentration of soluble PC-1 protein in patient plasma is used as a diagnostic. PC-1 circulates in human plasma and low plasma PC-1 level is independently associated with several features of the insulin resistant "metabolic syndrome" including abdominal fat distribution, high blood pressure and, with respect to lipid metabolism, insulin resistance.

PC-1 is a class II membrane glycoprotein that inhibits activation of insulin receptor tyrosine kinase, and is associated with insulin resistance. A novel polymorphism in exon 4 of the PC-1 gene is significantly correlated with insulin resistance. The subject genes and fragments thereof, encoded protein, and antibodies specific for the insulin resistance associated forms of PC-1 are useful in characterizing patients for susceptibility to insulin resistance. Such screening methods may be used in conjunction with counseling and preventive measures.

Nucleic Acid Compositions

As used herein, the term PC-1 genes and encoded polypeptides shall be used to generally designate any of the mammalian PC-1 genes and gene products, and unless otherwise stated will be the human homolog. The sequences of the invention comprise a sequence polymorphism, generally resulting in a change in coding sequence, that confer a susceptibility to insulin resistance, and may lead to hyperglycemia and NIDDM. Such polymorphisms may be generically referred to herein as a resistance associated PC-1 sequence, or PC-1^R. Counseling and preventive measures are particularly important for such patients, and early diagnosis provides information concerning such a predisposition.

The effect of a candidate sequence polymorphism on PC-1 expression or function may be determined by kindred analysis for segregation of the sequence variation with the disease phenotype, e.g. insulin resistance, hyperglycemia, etc. A predisposing mutation will segregate with incidence of the disease. The subject mutations generally have a dominant phenotype, where a single altered allele will confer disease susceptibility. The penetrance will vary with the specific mutation.

As an alternative to kindred studies, biochemical studies are performed to determine whether a candidate sequence variation in the PC-1 coding region or control regions affects the quantity or function of the protein. The effect of a sequence variation on the interaction between PC-1 and insulin receptor or other tyrosine kinases is determined by binding studies or kinase assays, where a decreased level of inhibition or binding is indicative of a predisposing mutation. Normal PC-1 will inhibit the tyrosine kinase activity of the insulin receptor, but not other tyrosine kinases.

In one embodiment of the invention, polymorphisms of interest provide for amino acid substitutions in the extracellular domain of PC-1, particularly the cysteine-rich domain, which may substitute a charged amino acid with a neutral amino acid. In one embodiment of the invention the amino acid substitution is at a lysine residue in this region, e.g. K121 or K119. Polymorphisms at these residues, where the lysine is substituted with any of the other 19 naturally occurring amino acids, may be referred to generically as a [*121] PC-1 or [*119] PC-1 polymorphisms. Specific polymorphisms of interest substitute a neutral amino acid in place of the lysine, particularly glutamine or arginine. A naturally occurring polymorphism associated with insulin resistance comprises a lys→glu substitution at position 121 of the protein, herein referred to as "[K121Q] PC-1", or merely "[Q] PC-1".

The human [Q] PC-1 amino acid sequence is provided as SEQ ID NO:4, and the encoding gene as SEQ ID NO:3. In order to identify the subject PC-1 polymorphisms, exonic primers from the published sequence data were used to isolate genomic clones. Sequence data from the genomic clones was used to generate specific primers. These primers

were used to amplify genomic DNA. The PCR products were screened for mutations using single strand conformation polymorphism (SSCP) analysis. The specific polymorphism found in SEQ ID NO:3 was identified in a number of patients.

DNA encoding a PC-1^R protein may be cDNA or genomic DNA or a fragment thereof that encompasses the altered residue, e.g. [*121] PC-1. As known in the art, cDNA sequences have the arrangement of exons found in processed mRNA, forming a continuous open reading frame, while genomic sequences may have introns interrupting the open reading frame. The term "[*121] PC-1 gene" shall be intended to mean the open reading frame encoding such specific PC-1 polypeptides, as well as adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression, up to about 1 kb beyond the coding region, in either direction. The intron-exon boundaries of the PC-1 gene are provided in the examples.

Genomic sequences of interest comprise the nucleic acids present between the initiation codon and the stop codon, including all of the introns that are normally present in a native chromosome. It may include the 3' and 5' untranslated regions found in the mature mRNA. It may further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, etc., including about 1 kb of flanking genomic DNA at either the 5' or 3' end of the coding region. The genomic DNA may be isolated as a fragment of 50 kbp or smaller; and substantially free of flanking chromosomal sequence.

The genomic PC-1 5' and 3' sequence, including specific transcriptional and translational regulatory sequences, such as promoters, enhancers, etc., including about 1 kb, but possibly more, of flanking genomic DNA at the 5' end of the transcribed region, is of particular interest. The promoter region is useful for determining the pattern of PC-1 expression, e.g. induction and inhibition of expression in various tissues, and for providing promoters that mimic these native patterns of expression. A polymorphic PC-1 regulatory sequence, i.e. including one or more of the provided 3' UTR polymorphisms, is useful for expression studies to determine the effect of sequence variations on mRNA expression and stability. The polymorphisms are also used as single nucleotide polymorphisms to detect genetic linkage to phenotypic variation in activity and expression of PC-1. The polymorphic 3' UTR sequences are provided as SEQ ID NO:6 ("A" allele); SEQ ID NO:7 ("P" allele); and SEQ ID NO:8 ("N" allele). The polymorphisms are as follows:

nucleotide position	127	136	178
SEQ ID NO: 6	G	G	C
SEQ ID NO: 7	A	C	T
SEQ ID NO: 8	A	G	T

The promoter region of PC-1 is provided as SEQ ID NO:5. The promoter region is useful for determining natural patterns of expression, particularly those that may be associated with disease. Alternatively, mutations may be introduced into the promoter region to determine the effect of altering expression in experimentally defined systems. The promoter also finds use in the construction of animal models where it is desirable to mimic the native patterns of PC-1 expression. Methods for the identification of specific DNA motifs involved in the binding of transcriptional factors are

known in the art, e.g. sequence similarity to known binding motifs, gel retardation studies, etc. For examples, see Blackwell et al. (1995) *Mol Med* 1: 194-205; Mortlock et al. (1996) *Genome Res.* 6: 327-33; and Joulin and Richard-Foy (1995) *Eur J Biochem* 232: 620-626. Specific regulatory motifs are found in the provided promoter sequence at positions: SEQ ID NO:5; 192-205; and SEQ ID NO:5, 453-458.

The nucleic acid compositions of the subject invention encode all or a part of the subject polypeptides. Fragments may be obtained of the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, etc. For the most part, DNA fragments will be at least about 25 nt in length, usually at least about 30 nt, more usually at least about 50 nt. For use in amplification reactions, such as PCR, a pair of primers will be used. The exact composition of the primer sequences is not critical to the invention, but for most applications the primers will hybridize to the subject sequence under stringent conditions, as known in the art. It is preferable to chose a pair of primers that will generate an amplification product of at least about 50 nt, preferably at least about 100 nt. Algorithms for the selection of primer sequences are generally known, and are available in commercial software packages. Amplification primers hybridize to complementary strands of DNA, and will prime towards each other. Amplification primers of interest include the intron sequences flanking each exon, as shown in the examples, which may lie immediately outside of the coding sequence, or may span the actual junction. Use of such primers allows specific amplification of the exon sequence from genomic DNA.

The subject PC-1^R genes and associated sequences are isolated and obtained in substantial purity, generally as other than an intact mammalian chromosome. Usually, the DNA will be obtained substantially free of other nucleic acid sequences that do not include a PC-1 sequence or fragment thereof, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", i.e. flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

PC-1 Polypeptides

The subject nucleic acid compositions may be employed for producing PC-1^R protein, or fragments thereof that encompass a polymorphisms of interest, e.g. [Q121] PC-1. For expression, an expression cassette may be employed, providing for a transcriptional and translational initiation region, which may be inducible or constitutive, the coding region under the transcriptional control of the transcriptional initiation region, and a transcriptional and translational termination region. Various transcriptional initiation regions may be employed which are functional in the expression host.

The peptide may be expressed in prokaryotes or eukaryotes in accordance with conventional ways, depending upon the purpose for expression. For large scale production of the protein, a unicellular organism or cells of a higher organism, e.g. eukaryotes such as vertebrates, particularly mammals, may be used as the expression host, such as *E. coli*, *B. subtilis*, *S. cerevisiae*, and the like. In many situations, it may be desirable to express the subject PC-1 gene in a mammalian host, whereby the PC-1 gene product will be glycosylated, and secreted.

With the availability of the protein in large amounts by employing an expression host, the protein may be isolated

and purified in accordance with conventional ways. A lysate may be prepared of the expression host and the lysate purified using HPLC, exclusion chromatography, gel electrophoresis, affinity chromatography, or other purification technique. The purified protein will generally be at least about 80% pure, preferably at least about 90% pure, and may be up to and including 100% pure. By pure is intended free of other proteins, as well as of cellular debris.

The polypeptide is used for the production of antibodies, where short fragments provide for antibodies specific for the particular polypeptide, and larger fragments allow for the production of antibodies over the surface of the protein. Antibodies may be raised to the normal or insulin resistant forms of PC-1. Of particular interest are antibodies that specifically recognize the insulin resistant forms of the protein, i.e. the antibodies do not bind to the normal form. Also of interest are antibodies that recognize the soluble forms of the protein. Antibodies may be raised to isolated peptides corresponding to these mutations, or to the native protein, e.g. by immunization with cells expressing PC-1, immunization with liposomes containing PC-1, etc. Such antibodies are useful in therapy and diagnosis.

Antibodies are prepared in accordance with conventional ways, where the expressed polypeptide or protein is used as an immunogen, by itself or conjugated to known immunogenic carriers, e.g. KLH, pre-S HBsAg, other viral or eukaryotic proteins, or the like. Various adjuvants may be employed, with a series of injections, as appropriate. For monoclonal antibodies, after one or more booster injections, the spleen is isolated, the splenocytes immortalized, and then screened for high affinity antibody binding. The immortalized cells, e.g. hybridomas, producing the desired antibodies may then be expanded. For further description, see *Monoclonal Antibodies: A Laboratory Manual*, Harlow and Lane eds., Cold Spring Harbor Laboratories, Cold Spring Harbor, N.Y., 1988. If desired, the mRNA encoding the heavy and light chains may be isolated and mutagenized by cloning in *E. coli*, and the heavy and light chains mixed to further enhance the affinity of the antibody. Alternatives to in vivo immunization as a method of raising antibodies include binding to phage display libraries, usually in conjunction with in vitro affinity maturation.

Phenotypic Indications

Insulin resistance is an essential feature of a great variety of clinical disorders in addition to diabetes, including coronary artery disease, hyperlipidemia, obesity and hypertension. Individuals with non-insulin dependent diabetes have insulin resistance in peripheral tissues. They have a subnormal glucose utilization in skeletal muscle, where glucose transport across the cell membrane of skeletal muscle is the rate limiting step in glucose metabolism. In adipose and muscle cells, insulin stimulates a rapid and dramatic increase in glucose uptake, primarily by promoting the redistribution of the GLUT4 glucose transporter from its intracellular storage site to the plasma membrane. Impaired glucose tolerance (IGT) is associated with a normal fasting blood glucose but an elevated postprandial blood sugar between 7.8 and 11 mmol/L (140 and 199 mg/dL). Some patients with IGT are hyperinsulinimic, and progress to NIDDM.

The response to insulin has been measured by a number of different methods, and insulin resistance has been quantified by a number of different indices. A variety of procedures have been developed to detect the presence of insulin resistance. Using any of these techniques, there is a wide

range of insulin sensitivity in normal individuals, some of whose values overlap with similar values in people with diabetes. Therefore, one cannot distinguish between nondiabetic and diabetic individuals on the basis of measures of insulin resistance.

The most widely accepted research method or 'gold standard' is the euglycemic insulin clamp technique. With this procedure, exogenous insulin is infused, so as to maintain a constant plasma insulin level above fasting, while glucose is fixed at a basal level by infusing glucose at varying rates. This glucose infusion is delivered via an indwelling catheter at a rate based on plasma glucose measurements every 5 min. When the plasma glucose level falls below basal, the glucose infusion rate is increased to return plasma glucose to basal levels and vice versa. The amount of glucose infused over time (M value) is an index of insulin action on glucose metabolism. The more glucose that has to be infused per unit time, then the more sensitive the patient is to insulin. Conversely, the insulin-resistant patient requires much less glucose to maintain basal plasma glucose levels. The effect of insulin on fuel metabolism can be assessed in the absence of the confounding effects of hypoglycemic counterregulation, endogenous insulin secretion, or variable levels of hyperglycemia, and multiple insulin actions can be assessed by using isotopes, including regulation of glucose uptake and production, inhibition of lipolysis, and changes in protein metabolism.

An alternative is the minimal model. With this procedure, glucose and insulin are sampled frequently from an indwelling catheter during an intravenous glucose tolerance test; the results are entered into a computer model, which generates a value that is an index of insulin sensitivity (called Si). The acute insulin release (AIR) in response to glucose is also determined by the test. This measure of insulin resistance correlates reasonably well with the euglycemic insulin clamp in nondiabetic subjects. Its accuracy deteriorates in diabetes because the immediate plasma insulin response to the glucose challenge is diminished. Therefore, additional maneuvers are needed to raise plasma insulin levels, such as giving tolbutamide or exogenous insulin in the course of the test.

The most practical way of assessing insulin resistance is the homeostasis model assessment (HOMAIR), involving fasting insulin and glucose levels. This value is calculated as fasting plasma insulin (μ /mL) \times fasting plasma glucose (mmol/L) / 22.5 (Matthews et al. (1985) *Diabetologia*. 28:412-9). The steady-state basal plasma glucose and insulin concentrations are determined by their interaction in a feedback loop. A computer-solved model is been used to predict the homeostatic concentrations which arise from varying degrees beta-cell deficiency and insulin resistance. Comparison of a patient's fasting values with the model's predictions allows a quantitative assessment of the contributions of insulin resistance and deficient beta-cell function to the fasting hyperglycaemia. The estimate of insulin resistance obtained by homeostasis model assessment correlates with estimates obtained by use of the euglycaemic clamp, the fasting insulin concentration, and the hyperglycaemic clamp. The lower limit of the top quintile of HOMA(IR) distribution (i.e. 2.77) in nonobese subjects with no metabolic disorders has been chosen as the threshold for insulin resistance in some studies (Bonora et al. (1998) *Diabetes* 47:1643-9). The results of this study documented that 1) in hypertriglyceridemia and a low HDL cholesterol state, insulin resistance is as common as in NIDDM, whereas it is less frequent in hypercholesterolemia, hyperuricemia, and hypertension; 2) the vast majority of subjects with multiple

metabolic disorders are insulin resistant; 3) in isolated hypercholesterolemia, hyperuricemia, or hypertension, insulin resistance is not more frequent than can be expected by chance alone; and 4) in the general population, insulin resistance can be found even in the absence of any major metabolic disorders.

The measurement of insulin concentration can be done in the overnight fasted condition, since in the postprandial state, glucose levels are changing rapidly and the variable levels of glucose confound the simultaneous measure of insulin levels as an index of insulin action. There is a significant correlation between fasting insulin levels and insulin action as measured by the clamp technique. Very high plasma insulin values in the setting of normal glucose levels are very likely to reflect insulin resistance. As individuals develop diabetes, plasma glucose increases and plasma insulin decreases and so the plasma insulin level no longer reflects only insulin resistance because it becomes influenced by the appearance of a β -cell defect and hyperglycemia.

Detection of PC-1 Associated Insulin Resistance

DNA from a patient having insulin resistance, as described above, suspected of association with aberrant PC-1 is analyzed for the presence of an IR polymorphism. Genetic characterization analyzes DNA or RNA, from any source, e.g. skin, cheek scrapings, blood samples, etc. The nucleic acids are screened for the presence of an insulin resistant polymorphism, e.g. SEQ ID NO:3, as compared to a normal sequence (SEQ ID NO:1, SEQ ID NO:2).

A number of methods are available for analyzing nucleic acids for the presence or absence of a specific sequence. Where large amounts of DNA are available, genomic DNA is used directly. Analysis of genomic DNA may use whole chromosomes or fractionated DNA, e.g. Southern blots, etc. Comparative Genomic Hybridization (CGH), as described in U.S. Pat. No. 5,665,549, provides methods for determining the relative number of copies of a genomic sequence. The intensity of the signals from each labeled subject nucleic acid and/or the differences in the ratios between different signals from the labeled subject nucleic acid sequences are compared to determine the relative copy numbers of the nucleic acid sequences as a function of position along the reference chromosome spread. Other methods for fluorescence in situ hybridization are known in the art, for a review, see Fox et al. (1995) *Clin Chem* 41(11):1554-1559.

Alternatively, the region of interest is cloned into a suitable vector and grown in sufficient quantity for analysis. Cells that express PC-1 may be used as a source of mRNA, which may be assayed directly or reverse transcribed into cDNA for analysis. The nucleic acid may be amplified by conventional techniques, such as the polymerase chain reaction (PCR), to provide sufficient amounts for analysis. The use of the polymerase chain reaction is described in Saiki, et al. (1985) *Science* 239:487, and a review of techniques may be found in Sambrook, et al. *Molecular Cloning: A Laboratory Manual*, CSH Press 1989, pp.14.2-14.33. Alternatively, various methods are known in the art that utilize oligonucleotide ligation as a means of detecting polymorphisms, for examples see Riley et al. (1990) *N.A.R.* 18:2887-2890; and Delahunty et al. (1996) *Am. J. Hum. Genet.* 58:1239-1246.

A detectable label may be included in an amplification reaction. Suitable labels include fluorochromes, e.g. fluorescein isothiocyanate (FITC), rhodamine, Texas Red,

phycoerythrin, allophycocyanin, 6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE), 6-carboxy-X-rhodamine (ROX), 6-carboxy-2',4',7',4',7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM) or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA), radioactive labels, e.g. ^{32}P , ^{35}S , ^3H ; etc. The label may be a two stage system, where the amplified DNA is conjugated to biotin, haptens, etc. having a high affinity binding partner, e.g. avidin, specific antibodies, etc., where the binding partner is conjugated to a detectable label. The label may be conjugated to one or both of the primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

The sample nucleic acid, e.g. genomic DNA, amplification product or cloned fragment, is analyzed by one of a number of methods known in the art. The nucleic acid may be sequenced by dideoxy or other methods, and the sequence of bases compared to a wild-type PC-1 sequence. Hybridization with the variant sequence may also be used to determine its presence, by Southern blots, dot blots, etc. The hybridization pattern of a control and variant sequence to an array of oligonucleotide probes immobilised on a solid support, as described in U.S. Pat. No. 5,445,934, or in WO95/35505, may also be used as a means of detecting the presence of variant sequences. Single strand conformational polymorphism (SSCP) analysis, denaturing gradient gel electrophoresis (DGGE), and heteroduplex analysis in gel matrices are used to detect conformational changes created by DNA sequence variation as alterations in electrophoretic mobility.

Alternatively, where a polymorphism creates or destroys a recognition site for a restriction endonuclease, the sample is digested with that endonuclease, and the products size fractionated to determine whether the fragment was digested. Fractionation is performed by gel or capillary electrophoresis, particularly acrylamide or agarose gels. The [Q] PC-1 allele has an Avail site that is not present in the [K] PC-1 allele, and this difference may be exploited for genetic screening.

Changes in the promoter or enhancer sequence that may affect expression levels of PC-1 can be compared to expression levels of the normal allele by various methods known in the art. Methods for determining promoter or enhancer strength include quantitation of the expressed natural protein; insertion of the variant control element into a vector with a reporter gene such as β -galactosidase, luciferase, chloramphenicol acetyltransferase, etc. that provides for convenient quantitation; and the like.

Diagnostic screening may also be performed for polymorphisms that are genetically linked to a predisposition for PC-1 associated insulin resistance, particularly through the use of microsatellite markers, e.g. the variable repeat in intron 3, or single nucleotide polymorphisms, e.g. the 3' UTR polymorphisms. Frequently the microsatellite polymorphism itself is not phenotypically expressed, but is linked to sequences that result in a disease predisposition. However, in some cases the microsatellite sequence itself may affect gene expression. Microsatellite linkage analysis may be performed alone, or in combination with direct detection of polymorphisms, as described above. The use of microsatellite markers for genotyping is well documented. For examples, see Mansfield et al. (1994) *Genomics* 24:225-233; Ziegler et al. (1992) *Genomics* 14:1026-1031; Dib et al., supra.

Microsatellite loci that are useful in the subject methods have the general formula:

$$U(R)_nU',$$

where U and U' are non-repetitive flanking sequences that uniquely identify the particular locus, R is a repeat motif, and n is the number of repeats. The repeat motif is at least 2 nucleotides in length, up to 7, usually 2-4 nucleotides in length. Repeats can be simple or complex. The flanking sequences U and U' uniquely identify the microsatellite locus within the human genome. U and U' are at least about 18 nucleotides in length, and may extend several hundred bases up to about 1 kb on either side of the repeat. Within U and U', sequences are selected for amplification primers. The exact composition of the primer sequences are not critical to the invention, but they must hybridize to the flanking sequences U and U', respectively, under stringent conditions. Criteria for selection of amplification primers are as previously discussed. To maximize the resolution of size differences at the locus, it is preferable to choose a primer sequence that is close to the repeat sequence, such that the total amplification product is between 100-500 nucleotides in length.

The number of repeats at a specific locus, n, is polymorphic in a population, thereby generating individual differences in the length of DNA that lies between the amplification primers. The number will vary from at least 1 repeat to as many as about 100 repeats or more.

The primers are used to amplify the region of genomic DNA that contains the repeats. Conveniently, a detectable label will be included in the amplification reaction, as previously described. Multiplex amplification may be performed in which several sets of primers are combined in the same reaction tube. This is particularly advantageous when limited amounts of sample DNA are available for analysis. Conveniently, each of the sets of primers is labeled with a different fluorochrome.

After amplification, the products are size fractionated. Fractionation may be performed by gel electrophoresis, particularly denaturing acrylamide or agarose gels. A convenient system uses denaturing polyacrylamide gels in combination with an automated DNA sequencer, see Hunkapillar et al. (1991) *Science* 254:59-74. The automated sequencer is particularly useful with multiplex amplification or pooled products of separate PCR reactions. Capillary electrophoresis may also be used for fractionation. A review of capillary electrophoresis may be found in Landers, et al. (1993) *BioTechniques* 14:98-111. The size of the amplification product is proportional to the number of repeats (n) that are present at the locus specified by the primers. The size will be polymorphic in the population, and is therefore an allelic marker for that locus.

Screening for polymorphisms in PC-1 may be based on the functional or antigenic characteristics of the protein. Protein truncation assays are useful in detecting deletions that may affect the biological activity of the protein. Various immunoassays designed to detect polymorphisms in PC-1 proteins may be used in screening. Where many diverse genetic mutations lead to a particular disease phenotype, functional protein assays have proven to be effective screening tools, for example by detecting the specific phosphatase activity on a PC-1 substrate. Alternatively, changes in electrophoretic mobility may be used.

Antibodies specific for an PC-1^R polymorphism may be used in staining or in immunoassays. Samples, as used herein, include cells, e.g. biopsy samples, biological fluids such as semen, blood, cerebrospinal fluid, tears, saliva,

lymph, dialysis fluid and the like; organ or tissue culture derived fluids; and fluids extracted from physiological tissues. Also included in the term are derivatives and fractions of such fluids. The cells may be dissociated, in the case of solid tissues, or tissue sections may be analyzed. Alternatively a lysate of the cells may be prepared.

Diagnosis may be performed by a number of methods to determine the absence or presence or altered amounts of normal or PC-1^R in patient cells. For example, detection may utilize staining of cells or histological sections, performed in accordance with conventional methods. Cells are permeabilized to stain cytoplasmic molecules. The antibodies of interest are added to the cell sample, and incubated for a period of time sufficient to allow binding to the epitope, usually at least about 10 minutes. The antibody may be labeled with radioisotopes, enzymes, fluorescers, chemiluminescers, or other labels for direct detection. Alternatively, a second stage antibody or reagent is used to amplify the signal. Such reagents are well known in the art. For example, the primary antibody may be conjugated to biotin, with horseradish peroxidase-conjugated avidin added as a second stage reagent. Alternatively, the secondary antibody conjugated to a fluorescent compound, e.g. fluorescein, rhodamine, Texas red, etc. Final detection uses a substrate that undergoes a color change in the presence of the peroxidase. The absence or presence of antibody binding may be determined by various methods, including flow cytometry of dissociated cells, microscopy, radiography, scintillation counting, etc.

An alternative method for diagnosis depends on the in vitro detection of binding between antibodies and polymorphic PC-1^R in a lysate. Measuring the concentration of PC-1^R binding in a sample or fraction thereof may be accomplished by a variety of specific assays. A conventional sandwich type assay may be used. For example, a sandwich assay may first attach PC-1^R specific antibodies to an insoluble surface or support. Patient sample lysates are then added to the supports (for example, separate wells of a microtiter plate) containing antibodies. Preferably, a series of standards, containing known concentrations of normal and/or PC-1^R is assayed in parallel with the samples or aliquots thereof to serve as controls. The quantitation may then be performed by adding a labeled antibody specific for PC-1^R. Other immunoassays are known in the art and may find use as diagnostics. Ouchterlony plates provide a simple determination of antibody binding. Western blots may be performed on protein gels or protein spots on filters, using a detection system specific for PC-1 as desired, conveniently using a labeling method as described for the sandwich assay.

Immunoassays may also be used in the detection of soluble PC-1 in the plasma of a patient, where quantitative and qualitative analysis may be performed. It is found that decreased levels of PC-1 in the plasma are associated with increased levels in the muscle, therefore a relatively low titer is associated with insulin resistance. In addition, the soluble PC-1 may be analyzed for the presence of a predisposing polymorphism, e.g. that Q121 protein.

A kit may be provided for practice of the subject diagnostic methods. Such a kit may contain hybridization probes that bind to a polymorphic PC-1^R sequence under hybridization conditions where the probe does not bind to a wild type PC-1 sequence. Alternatively, antibodies specific for a polymorphic PC-1^R polypeptide may be included. Such a kit will typically include positive and negative nucleic acid or polypeptide controls, and such other buffers and reagents as may be necessary to practice the method.

Modulation of Gene Expression

The PC-1 genes, gene fragments, or the encoded protein or protein fragments are useful in gene therapy to treat

disorders associated with PC-1 insulin resistance. Expression vectors may be used to introduce a PC-1 gene into a cell. Such vectors generally have convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences. Transcription cassettes may be prepared comprising a transcription initiation region, the target gene or fragment thereof, and a transcriptional termination region. The transcription cassettes may be introduced into a variety of vectors, e.g. plasmid; retrovirus, e.g. lentivirus; adenovirus; and the like, where the vectors are able to transiently or stably be maintained in the cells, usually for a period of at least about one day, more usually for a period of at least about several days to several weeks.

The gene or PC-1 protein may be introduced into tissues or host cells by any number of routes, including viral infection, microinjection, or fusion of vesicles. Jet injection may also be used for intramuscular administration, as described by Furth et al. (1992) *Anal Biochem* 205:365-368. The DNA may be coated onto gold microparticles, and delivered intradermally by a particle bombardment device, or "gene gun" as described in the literature (see, for example, Tang et al. (1992) *Nature* 356:152-154), where gold microparticles are coated with PC-1 protein or nucleic acids encoding PC-1, then bombarded into skin cells.

Antisense molecules can be used to down-regulate expression of PC-1 in cells. The anti-sense reagent may be antisense oligonucleotides (ODN), particularly synthetic ODN having chemical modifications from native nucleic acids, or nucleic acid constructs that express such anti-sense molecules as RNA. The antisense sequence is complementary to the mRNA of the targeted gene, and inhibits expression of the targeted gene products. Antisense molecules inhibit gene expression through various mechanisms, e.g. by reducing the amount of mRNA available for translation, through activation of RNase H, or steric hindrance. One or a combination of antisense molecules may be administered, where a combination may comprise multiple different sequences.

Alternatively, the antisense molecule is a synthetic oligonucleotide. Antisense oligonucleotides will generally be at least about 7, usually at least about 12, more usually at least about 20 nucleotides in length, and not more than about 500, usually not more than about 50, more usually not more than about 35 nucleotides in length, where the length is governed by efficiency of inhibition, specificity, including absence of cross-reactivity, and the like. It has been found that short oligonucleotides, of from 7 to 8 bases in length, can be strong and selective inhibitors of gene expression (see Wagner et al. (1996) *Nature Biotechnology* 14:840-844).

A specific region or regions of the endogenous sense strand mRNA sequence is chosen to be complemented by the antisense sequence, preferably encompassing the [Q121] PC-1 mutation. Selection of a specific sequence for the oligonucleotide may use an empirical method, where several candidate sequences are assayed for inhibition of expression of the target gene in an in vitro or animal model. A combination of sequences may also be used, where several regions of the mRNA sequence are selected for antisense complementation.

Nucleic acids may be naturally occurring, e.g. DNA or RNA, or may be synthetic analogs, as known in the art. Such analogs may be preferred for use as probes because of superior stability under assay conditions. Modifications in the native structure, including alterations in the backbone, sugars or heterocyclic bases, have been shown to increase

intracellular stability and binding affinity. Among useful changes in the backbone chemistry are phosphorothioates; phosphorodithioates, where both of the non-bridging oxygens are substituted with sulfur; phosphoroamidites; alkyl phosphotriesters and boranophosphates. Achiral phosphate derivatives include 3'-O'-5'-S-phosphorothioate, 3'-S-5'-O-phosphorothioate, 3'-CH₂-5'-O-phosphonate and 3'-NH-5'-O-phosphoroamidate. Peptide nucleic acids replace the entire ribose phosphodiester backbone with a peptide linkage.

Sugar modifications are also used to enhance stability and affinity. The α -anomer of deoxyribose may be used, where the base is inverted with respect to the natural β -anomer. The 2'-OH of the ribose sugar may be altered to form 2'-O-methyl or 2'-O-allyl sugars, which provides resistance to degradation without comprising affinity.

Modification of the heterocyclic bases must maintain proper base pairing. Some useful substitutions include deoxyuridine for deoxythymidine; 5-methyl-2'-deoxycytidine and 5-bromo-2'-deoxycytidine for deoxycytidine. 5-propynyl-2'-deoxyuridine and 5-propynyl-2'-deoxycytidine have been shown to increase affinity and biological activity when substituted for deoxythymidine and deoxycytidine, respectively.

The antisense molecules and/or other inhibitory agents are administered by contact with cells under conditions that permit entry. The molecules may be provided in solution or in any other pharmacologically suitable form for administration, such as a liposome suspension. There are many delivery methods known in the art for enhancing the uptake of nucleic acids by cells. Useful delivery systems include Sendai virus-liposome delivery systems (see Rapaport and Shai (1994) *J. Biol. Chem.* 269:15124-15131), cationic liposomes, polymeric delivery gels or matrices, porous balloon catheters (as disclosed by Shi et al. (1994) *Circulation* 90:955-951; and Shi et al. (1994) *Gene Therapy* 1:408-414), retrovirus expression vectors, and the like.

The use of liposomes as a delivery vehicle is one method of interest. The liposomes fuse with the cells of the target site and deliver the contents of the lumen intracellularly. The liposomes are maintained in contact with the cells for sufficient time for fusion, using various means to maintain contact, such as isolation, binding agents, and the like. Liposomes may be prepared with purified proteins or peptides that mediate fusion of membranes, such as Sendai virus or influenza virus, etc. The lipids may be any useful combination of known liposome forming lipids, including cationic lipids, such as phosphatidylcholine. The remaining lipid will normally be neutral lipids, such as cholesterol, phosphatidyl serine, phosphatidyl glycerol, and the like.

The therapeutic agents are administered at a dose effective to reduce expression level of PC-1^R at least about 50%, more usually at least 80%, and preferably to substantially undetectable levels.

Genetically Modified Cells and Animals

The subject nucleic acids can be used to generate transgenic animals or site specific gene modifications in cell lines. Transgenic animals may be made through homologous recombination. Alternatively, a nucleic acid construct is randomly integrated into the genome. Vectors for stable integration include plasmids, retroviruses and other animal viruses, YACs, and the like. The modified cells or animals are useful in the study of PC-1 function and regulation. A detectable marker, such as lac Z may be introduced into the PC-1 locus, where upregulation of PC-1 expression will result in an easily detected change in phenotype.

DNA constructs for homologous recombination will comprise at least a portion of a polymorphic PC-1^R gene with the desired genetic modification, and will include regions of homology to the target locus. DNA constructs for random integration need not include regions of homology to mediate recombination. Conveniently, markers for positive and negative selection are included. Methods for generating cells having targeted gene modifications through homologous recombination are known in the art. For various techniques for transfecting mammalian cells, see Keown et al. (1990) *Methods in Enzymology* 185:527-537.

For embryonic stem (ES) cells, an ES cell line may be employed, or ES cells may be obtained freshly from a host, e.g. mouse, rat, guinea pig, etc. Such cells are grown on an appropriate fibroblast-feeder layer or grown in the presence of leukemia inhibiting factor (LIF). When ES cells have been transformed, they may be used to produce transgenic animals. After transformation, the cells are plated onto a feeder layer in an appropriate medium. Cells containing the construct may be detected by employing a selective medium. After sufficient time for colonies to grow, they are picked and analyzed for the occurrence of homologous recombination or integration of the construct. Those colonies that are positive may then be used for embryo manipulation and blastocyst injection. Blastocysts are obtained from 4 to 6 week old superovulated females. The ES cells are trypsinized, and the modified cells are injected into the blastocoel of the blastocyst. After injection, the blastocysts are returned to each uterine horn of pseudopregnant females. Females are then allowed to go to term and the resulting litters screened for mutant cells having the construct. By providing for a different phenotype of the blastocyst and the ES cells, chimeric progeny can be readily detected.

The chimeric animals are screened for the presence of the modified gene and males and females having the modification are mated to produce homozygous progeny. The transgenic animals may be any non-human mammal, such as laboratory animals, domestic animals, etc. The transgenic animals may be used in functional studies, drug screening, etc., e.g. to determine the effect of a candidate drug on insulin resistance.

Drug Screening Assays

Drug screening identifies agents inhibit or otherwise modulate PC-1 function in cells. Of particular interest are screening assays for agents that have a low toxicity for human cells. A wide variety of assays may be used for this purpose, including labeled in vitro protein-protein binding assays, electrophoretic mobility shift assays, immunoassays for protein binding, and the like. The purified protein may also be used for determination of three-dimensional crystal structure, which can be used for modeling intermolecular interactions, transporter function, etc.

The term "agent" as used herein describes any molecule, e.g. protein or pharmaceutical, with the capability of altering or mimicking the physiological function of PC-1. Generally a plurality of assay mixtures are run in parallel with different agent concentrations to obtain a differential response to the various concentrations. Typically, one of these concentrations serves as a negative control, i.e. at zero concentration or below the level of detection.

Candidate agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 50 and less than about 2,500 daltons. Candidate agents comprise functional groups necessary for structural

interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The candidate agents often comprise cyclized carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Candidate agents are also found among biomolecules including peptides, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof.

Candidate agents are obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides and oligopeptides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means, and may be used to produce combinatorial libraries. Known pharmacological agents may be subjected to directed or random chemical modifications, such as acylation, alkylation, esterification, amidification, etc. to produce structural analogs.

Where the screening assay is a binding assay, one or more of the molecules may be joined to a label, where the label can directly or indirectly provide a detectable signal. Various labels include radioisotopes, fluorescers, chemiluminescers, enzymes, specific binding molecules, particles, e.g. magnetic particles, and the like. Specific binding molecules include pairs, such as biotin and streptavidin, digoxin and antidigoxin etc. For the specific binding members, the complementary member would normally be labeled with a molecule that provides for detection, in accordance with known procedures.

A variety of other reagents may be included in the screening assay. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc that are used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Reagents that improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc. may be used. The mixture of components are added in any order that provides for the requisite binding. Incubations are performed at any suitable temperature, typically between 4 and 40° C. Incubation periods are selected for optimum activity, but may also be optimized to facilitate rapid high-throughput screening. Typically between 0.1 and 1 hours will be sufficient.

The compounds having the desired pharmacological activity may be administered in a physiologically acceptable carrier to a host for treatment of insulin resistance or hyperglycemia attributable to a defect in PC-1 function. The compounds may also be used to inhibit PC-1 function in resistance to insulin, etc. The inhibitory agents may be administered in a variety of ways, orally, topically, parenterally e.g. subcutaneously, intraperitoneally, by viral infection, intravascularly, etc. Topical treatments are of particular interest. Depending upon the manner of introduction, the compounds may be formulated in a variety of ways. The concentration of therapeutically active compound in the formulation may vary from about 0.1-100 wt. %.

The pharmaceutical compositions can be prepared in various forms, such as granules, tablets, pills, suppositories,

capsules, suspensions, salves, lotions and the like. Pharmaceutical grade organic or inorganic carriers and/or diluents suitable for oral and topical use can be used to make up compositions containing the therapeutically-active compounds. Diluents known to the art include aqueous media, vegetable and animal oils and fats. Stabilizing agents, wetting and emulsifying agents, salts for varying the osmotic pressure or buffers for securing an adequate pH value, and skin penetration enhancers can be used as auxiliary agents.

The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the subject invention, and are not intended to limit the scope of what is regarded as the invention. Efforts have been made to ensure accuracy with respect to the numbers used (e.g. amounts, temperature, concentrations, etc.) but some experimental errors and deviations should be allowed for. Unless otherwise indicated, parts are parts by weight, molecular weight is average molecular weight, temperature is in degrees centigrade; and pressure is at or near atmospheric.

EXPERIMENTAL

EXAMPLE 1

Polymorphic Variant of PC-1 Associated with Insulin Resistance

Methods

Subjects

127 unrelated, healthy, non obese subjects (body mass index, BMI, <30 Kg/m²) normotensive (blood pressure<140/90 mm Hg), normal glucose tolerant (by OGTT) were studied. Plasma insulin levels were measured before and during an OGTT that was carried out after 8 days on a weight-maintaining diet. Insulin stimulated glucose disposal was carried out in a subgroup of 71 subjects by the euglycaemic, hyperinsulinemic clamp.

Also studied were 132 type 2 diabetic patients (age=66.5±8.0 yr, 60 male/72 female, BMI=28.9±4.5 Kg/m²) with a strong family history of diabetes (one first degree relative with type 2 diabetes at the minimum). To minimize the possible inclusion of individuals affected by late onset type 1 diabetes, patients were selected on the basis of age of diabetes onset≥45 yrs, BMI≥21 Kg/m² and no need for insulin therapy.

Informed consent was obtained from participants before entry into the study, which was approved by the local research ethics committee.

Polymorphism Screening

Overlapping cosmid clones containing the PC-1 gene were isolated by screening a human chromosome 6 specific genomic library with human full length PC-1 cDNA. Cosmids were digested with different four base cutter restriction enzymes, blotted and hybridized to oligonucleotides designed on the cDNA sequence. Positive fragments were cloned and automatically sequenced. Intron-exon junctions were deduced comparing genomic and cDNA sequences.

All exon amplimers, obtained using specific oligonucleotides as primers, were analyzed in 40 unrelated and unselected individuals by Single Strand Conformation Polymorphism (SSCP) which was performed as follows. Amplification reaction products were denatured for 5 minutes at 87° C. in 90% formamide, 20 mM EDTA, 10 mM NaOH. After denaturation samples were chilled on ice, loaded on a native 8%-12% (according to amplimer size) acrylamide (29:1 Acrylamide-Bisacrylamide) gel (0.04×20×42 cm) in TBE and electrophoresed at 10 W constant power

for 13-16 hours at room temperature. After the electrophoresis, gels were treated by silver staining. PCR products showing different migration patterns at SSCP were cloned in a TA-cloning vector (Stratagene) and four clones for each sample were automatically sequenced from both ends.

Exon 4 amplimers were obtained using oligonucleotides 4Fw [SEQ ID NO:9] (5'-ctgtgttcactttggacatgttg-3') and 4Rv [SEQ ID NO:10] (5'gacgttggagaataccaggttg-3') as primers. PCR products were digested using Avail restriction enzyme and run on 12% native polyacrylamide gel for 2 hours at 500V. After the electrophoresis, gels were stained by silver nitrate. On the gel K alleles are displayed as single, uncut, bands of 238 bp, while Q alleles are shown as a doublet of 148 and 90 bp.

One-hundred-sixty unscreened blood donors were genotyped as background population. All genotypings were performed in duplicate for each individual and to prevent observer bias the investigator was unaware of sample origin. Skin Fibroblast Culture and Insulin Receptor Autophosphorylation

Fibroblast cultures were established from 4-mm forearm skin-punch biopsies. I¹²⁵ insulin binding data were obtained by inhibition-competition studies. IR-TK (receptor autophosphorylation) was determined exposing cells for 10 min to increasing insulin (0-100 nM) concentrations. Cells were then solubilized in 50 mM Hepes buffer, pH 7.6, containing 1% Triton X-100, 1 mM PMSF, 2 mM orthovanadate and 1% BSA and IR-TK determined.

PC-1 Content in Muscle Tissue Specimens

Muscle tissue specimens were obtained from the external oblique muscle at elective abdominal surgery (cholecystectomy) and were immediately frozen in liquid nitrogen. Soluble extracts were prepared from frozen muscle tissue and PC-1 content was measured by a specific ELISA as previously described and normalized for protein content. Statistical Analysis

Group values are given as mean±SD. Student's t-test or Mann Whitney U test were used to compare mean values of 2 groups. One-way ANOVA and both Student-Newman-Keuls and Bonferroni t-test were used to compare mean values of more than 2 groups. Two-way ANOVA test was used to compare insulin dose-response curves of IR-TK. Chi-square test was used to compare allele frequency.

Results

The PC-1 gene has been located on chromosome 6q22-23. Analysis of a YAC contig from the region, allowed it to be more finely mapped, to between markers D6S457 and WI-3398. Only exon 4, which extends from nucleotide 447 to 571 of the cDNA and codes for an extracellular portion of PC-1, showed a polymorphic variant. When screened by SSCP analysis and sequencing it revealed a frequent first position A→C transversion at codon 121 (considering the second in frame ATG as the start codon) (FIG. 1a). This single base change substitutes a glutamine for a lysine in a cysteine-rich region of PC-1 (SEQ ID NO:1 and SEQ ID NO:3, respectively), and creates an Avail restriction enzyme recognition site. Avail digestion of exon 4 amplimers cuts the Q allele PCR fragments, leaving the K allele undigested (FIG. 1b). In 160 uncharacterized blood donors, the Q allele frequency was 17.5%, with only 2 QQ homozygotes. The observed genotype frequencies were in agreement with those predicted by the Hardy-Weimberg equilibrium.

Having identified a PC-1 polymorphism which changes both amino acid composition and electric charge, and thus with potential biological relevance, we searched for an

association with insulin resistance. Accordingly we studied 127 unrelated, healthy, non obese, normotensive, non diabetic subjects resident in Sicily. As expected, these individuals showed a wide range of plasma insulin levels during OGTT, a finding which in the presence of normal glucose tolerance, indicates a wide range of insulin sensitivity. These data were confirmed by the euglycemic hyperinsulinemic glucose clamp, a more quantitative technique for the measurement of insulin sensitivity. In a subgroup of 71 individuals the M values for insulin stimulated glucose disposal ranged from 2.34 to 9.62 mg/Kg/min.

Table 1 summarizes the clinical features of these 2 groups. Q allele carriers showed higher fasting plasma glucose ($p < 0.001$) (Table 1 and FIG. 2a) values. They also showed higher plasma insulin values at 60 ($p < 0.05$) and 120 ($p < 0.01$) minutes during OGTT (FIGS. 2a and 2b).

TABLE 1

Clinical Characteristics of the subject studied					
Genotype	Gender (M/F)	Age (years)	BMI (Kg/m ²)	FPG (mmol/l)	FIRI (pmol/l)
KK (n = 45)	27/18	36.6 ± 2.1	23.8 ± 0.5	4.7 ± 0.1	49.0 ± 4.0
KQ or QQ (n = 22)	18/4	40.3 ± 3.1	24.2 ± 0.8	5.1 ± 0.1*	60.0 ± 8.0

Data are expressed as mean ± SEM.

*p < 0.01 vs. KK subjects

BMI = body mass index

FPG = fasting plasma glucose

FIRI = fasting immunoreactive insulin

In the subjects studied by glucose clamp, insulin stimulated glucose disposal was lower in Q allele carriers when compared to KK allele age, sex and BMI matched subjects. No difference was observed in insulin levels at steady state during clamp studies in the 2 groups (485±165 pmol/l vs 460±78). On the average, therefore, Q allele carriers were insulin resistant and maintained normal glucose tolerance due to compensatory hyperinsulinemia. Mean blood pressure, plasma total cholesterol, HDL cholesterol and triglyceride levels were not different between the 2 groups.

Of the 2 subjects with QQ alleles, one was a 35 yr. old male who was studied by euglycemic clamp and had the second lowest M value ($M = 2.57$ mg/Kg/min) of the all the XY males studied. The BMI (28 Kg/m²), blood pressure (138/90 mm Hg), and lipid profile (cholesterol/HDL ratio being 0.16 and triglycerides 176 mg/dl) were in the upper range of the studied individuals. The second QQ subject was a 52 yr. old female with BMI (21 Kg/m²) blood pressure, and her lipid profile was entirely normal. She did not agree to be studied by euglycemic clamp. Both QQ subjects were first degree relatives of a type 2 diabetic patient.

When subjects were subdivided into tertiles according to plasma insulin levels at 120 minutes during the OGTT (tertile 1=low, tertile 2=intermediate, and tertile=3 high insulin levels). As expected, the mean M value for glucose disposal progressively decreased from tertile 1 to tertile 3 (7.22 ± 0.26 mg/Kg/min, $n=21$ vs. 5.86 ± 0.28 , $n=25$ and 4.89 ± 0.23 , $n=25$, $p < 0.001$). Q allele frequency was similar in subjects from tertiles 1 and 2 (11.7, $n=34$ and 10.6%, $n=33$, respectively), but it was much higher in tertile 3 insulin resistant subjects (29.4%, $n=34$, $p < 0.01$ when compared to the remaining 67 subjects, 11.2%). Also, in 133 type 2 diabetic patients Q allele frequency was higher (20.8%, $p < 0.01$) than in tertile 1 and 2 subjects with no difference between obese ($BMI > 30$ Kg/m², $n=90$) and non obese ($n=42$) patients (21.1% and 20.2%, respectively).

TABLE 2

Q allele frequency and insulin sensitivity (M value) in subjects divided in tertiles according to plasma insulin level at 120 min during OGTT (IRI 120 min)

Tertiles IRI 120 min range	1 (1353-300 pmole/l)	2 (273-147 pmol/l)	3 (140-27 pmol/l)
	29# n = 22	15 n = 23	9 n = 22
Q allele frequency %			
M value	$4.50 \pm 0.36^*$	$5.46 \pm 0.30^{**}$	7.21 ± 0.26
(mg/kg/min)	n = 17	n = 20	n = 17

Data are expressed as mean ± SEM. Number of subjects are given in parenthesis.

#p < 0.05 vs tertile 3

*p < 0.01 vs. tertile 2 and 3

**p < 0.01 vs. tertile 3

In order to exclude any association of the Q allele variant with other changes in PC-1, each of the 25 exons from a QQ control were sequenced from the start to the stop codon. No other base change was detected.

In order to study IR autophosphorylation activity, cultured fibroblasts from 5 Q/K and 5 gender, age and BMI matched KK subjects were selected on the basis of a similar PC-1 protein content (50.3 ± 8.7 and 60.8 ± 15.4 ng/0.1 mg protein, respectively). Q/K fibroblasts showed a reduced IR autophosphorylation activity ($p < 0.01$) (FIG. 3). Insulin binding to its receptor was studied and no difference in both total specific binding (% of bound/total radioactivity= 0.52 ± 0.10 per 0.1 mg protein and 0.55 ± 0.11 in Q/K and KK subjects, respectively), and IC_{50} (0.27 ± 0.05 nmol/l and 0.26 ± 0.08).

PC-1 content was not significantly different in muscle specimens from 8 QK and 26 KK sex, age and BMI matched subjects (36.5 ± 5.1 ng/mg protein vs. 25.9 ± 2.6 in QK and KK subjects, respectively).

Discussion

The data provided herein demonstrate that a PC-1 gene polymorphism (K121Q in exon 4) is associated with decreased insulin sensitivity in healthy non-diabetic individuals. Because insulin resistance is a major risk factor for the development of type 2 diabetes, Q allele carriers may be at higher risk to develop diabetes. This is supported by the high Q allele frequency observed in patients with type 2 diabetes mellitus. No association was observed with BMI both in healthy and diabetic individuals.

We previously reported that increased PC-1 content in skeletal muscle and adipose tissue is associated with insulin resistance. In addition, when cultured cells overexpress PC-1 they are insulin resistant secondary to both decreased IR tyrosine kinase activity and reduced downstream signaling steps. These latter observations indicate that an increased PC-1 content may play a role in insulin resistance through the inhibition of IR-TK activity.

PC-1 content is not significantly different in skeletal muscle from KQ with respect to KK subjects, indicating that insulin resistance in Q allele subjects is not due to an increased PC-1 protein content. Again, these data suggest that structural differences between the 2 variant proteins may account for different insulin sensitivity, independent of protein content. These data indicate PC-1 is an important candidate for the genetic regulation of whole body insulin sensitivity. PC-1 genotyping can be used for identifying individuals at risk of developing insulin resistance.

EXAMPLE 2

Fasting Plasma PC-1 and its Regulation by Insulin

A soluble form of PC-1 is generated by intracellular cleavage of its transmembrane domain, and subsequently

released by the cell. It is not known whether soluble PC-1 circulates in human plasma. The possibility of measuring PC-1 in human plasma would considerably increase the feasibility of screening studies.

A sensitive and specific ELISA was set up, and used to measure plasma PC-1 concentration before and after a 2-hour euglycemic hyperinsulinemic clamp in 22 healthy control, and 27 subjects affected by diseases known to be associated with insulin resistance (i.e. obesity and essential hypertension). The obtained results indicate that low fasting level and abnormal acute regulation by insulin of plasma PC-1 concentration are associated with several features of the "metabolic syndrome", including abdominal fat distribution, high blood pressure and low insulin sensitivity on both glucose and lipid metabolism.

Methods

Plasma PC-1 Measurement

Wells in Maxisorb plates were precoated overnight incubation at 40° C. with an affinity purified polyclonal antibody to PC-1. After washing with TBST buffer (20 mM Tris, 150 mM NaCl, 0.05% Tween-20) to remove unbound antibody, wells were blocked with 150 μ l TBST containing 1% bovine serum albumin (BSA) (30 min at 56° C.), and washed again with TBST. Then, human plasma (10–30 μ l diluted to a total volume of 100 μ l with 50 mM HEPES buffer, pH 7.6, containing 0.05% Tween-20, 1 mM PMSF, 2 mM orthovanadate, 1% BSA and 1 mg/ml bacitracin) was added to each well and PC-1 was allowed to bind overnight at 4° C. After extensive washing with TBST, a biotinylated anti-PC-1 monoclonal antibody was added in the 50 mM HEPES buffer. After 2 hr at 22° C., peroxidase-streptavidin was added and 30 min later, wells were washed again with TBST and then 100 μ l of biotinyl-tyramide solution was added. After 15 min incubation at 22° C., wells were washed with TBST and streptavidin-horseradish peroxidase was added (30 min at 22° C.). After further extensive washing, the peroxidase activity was determined calorimetrically by adding 3,3',5,5'-tetramethylbenzidine (TMB) at a concentration of 0.4 g/l in an organic base, and measuring the absorbance at 451 nm.

Muscle PC-1 Measurement

Muscle tissue specimens were obtained from the external oblique muscle at elective abdominal surgery (cholecystectomy). After adipose tissue was dissected and blood removed, specimens were immediately frozen in liquid nitrogen. Soluble extracts were prepared from frozen muscle tissue as previously described. Briefly, muscle tissue (approximately 150 mg) was pulverized under liquid nitrogen and then homogenized in 2 ml buffer (50 mM HEPES, 150 mM NaCl, 2 mM PMSF, pH 7.6) at 4° C. using a polytron homogenizer for 10 sec. at medium speed. Triton X-100 was added to a final concentration of 1%, and the homogenates solubilized for 60 min at 4° C. The material was centrifuged at 100K g for 60 min at 4° C. and the supernatants used for the PC-1 content measurement.

Statistical Analysis

One way analysis of variance (ANOVA) was utilized when means values from 3 groups were compared. Paired Student's t test was utilized to compare mean values before and after clamp.

Correlation (either "Pearson" if the data was distributed normally or "Spearman" if the data was not distributed normally) analysis was used to look for numerical relationship between values. Statistically significant correlations were confirmed by linear regression analysis. Stepwise regression analysis was utilized for multiple correlations. Data are given as mean \pm SEM.

Results

Subjects Studied

Twenty two healthy control and 27 subjects affected by either obesity (BMI>28, n=10) or essential hypertension (mean blood pressure>108 mm Hg, n=12) or both (n=5) were studied. Clinical and metabolic features of the 49 subjects are shown in Table 3. As expected, insulin sensitivity, as indicated by M values derived by euglycemic hyperinsulinemic clamp studies, was significantly reduced in obese and/or hypertensive patients as compared to normal controls.

TABLE 3

	age	sex	BMI	W/HI	MBP	BG	IRI	M
<u>Control</u>								
mean	37	12/10	23.8	0.83	90	5.1	65	6.2
SEM	2		0.4	0.03	2	0.1	7	0.4
<u>Insulin Resistant</u>								
mean	47	19/8	29.2	0.92	109	5.3	80	4.8
SEM	2		0.8	0.02	3	0.1	7	0.3

Plasma PC-1 Concentration

Fasting plasma PC-1 was measured by ELISA. Human plasma produced a dilution slope that paralleled the PC-1 standard. Intra- and inter-assay coefficient of variations were <8%. Plasma PC-1 concentration ranged from 1 to 70 ng/ml with a mean \pm SE of 26.5 \pm 2.9 and a median of 24.5. No significant difference was observed between plasma PC-1 concentration in control (27.7 \pm 4.5, n=22) and insulin resistant obese and/or hypertensive (25.6 \pm 3.9, n=27) subjects.

When the 49 subjects were considered together, plasma PC-1 concentration was correlated negatively with both waist/hip ratio (-0.49 , $p=0.001$) and systolic blood pressure (-0.36 , $p=0.016$) and positively (0.40 , $p=0.01$) with the ability of insulin to suppress plasma FFA (delta FFA, calculated by subtracting basal FFA from FFA after the two hour euglycemic hyperinsulinemic clamp). Plasma PC-1 concentration remained significantly correlated with the waist/hip ratio also when data were adjusted for BMI and sex ($p=0.0024$), with systolic blood pressure also when data were adjusted for sex and age ($p=0.019$) and with delta FFA also when data were adjusted for BMI, sex and waist/hip ratio ($p=0.037$).

These data demonstrate that PC-1 circulates in human plasma and that low plasma PC-1 level is independently associated with several features of the "metabolic syndrome" including abdominal fat distribution, high blood pressure and, so far as lipid metabolism is concerned, insulin resistance.

Insulin Stimulated Values

In order to verify whether insulin exerts any effect on plasma PC-1, PC-1 was measured after two hour euglycemic hyperinsulinemic clamp. Although the mean plasma PC-1 concentration in the 49 subject after clamp was not different as compared to basal plasma PC-1 level (26.3 \pm 3.9, vs. 26.5 \pm 4.1), a wide range of the individual effects of insulin infusion were observed, from subjects showing an increase to subjects showing either no change or a reduction in plasma PC-1. When subjects were divided in tertiles according to their whole body insulin sensitivity on glucose values (M values), with the most insulin sensitive in tertile 1 and the most resistant in tertile 3, insulin stimulated PC-1 concentrations were significantly higher than basal plasma PC-1 concentration in subjects from tertile 1 (21.7 \pm 5.4 vs.

23

25.8±5.5 before and after clamp, respectively, $p=0.015$, $n=16$) but not in subjects from tertile 2 (31.7 ± 5.0 vs. 31.9 ± 4.6 , n.s., $n=17$) and 3 (25.9 ± 4.7 vs. 20.7 ± 3.7 , $n=16$). Moreover, the net effect of insulin on plasma PC-1 concentration (delta PC-1, calculated by subtracting basal PC-1 from insulin stimulated PC-1) was positively correlated with M value (0.37 , $p=0.009$) and negatively with BMI (-0.37 , $p=0.009$). A significant ($p=0.005$) positive correlation between M and delta PC-1 values was observed also when data was adjusted for BMI. A similar correlation between M and delta PC-1 values was observed also when control (0.049 , $p=0.05$, $n=22$) and insulin resistant obese and/or hypertensive (0.48 , $p=0.01$, $n=27$) subjects were considered separately).

These data demonstrate that insulin infusion is able to increase plasma PC-1 concentration in the most insulin sensitive subjects and that this effect is blunted in subjects with lower insulin sensitivity.

Plasma v. Muscle Tissue PC-1

In order to verify whether plasma PC-1 concentration was related to PC-1 content in skeletal muscle, we quantified PC-1 in both plasma and biopsied external oblique muscle of 9 additional subjects. PC-1 concentration in plasma was inversely correlated with PC-1 content in muscle (-0.9 ,

Intron Exon borders

[SEQ ID NO:11] Exon 1

CTCTCGCTG-GTAGGTCCGCGGCCAGGCCCGGCGCCCGGGAGGGCTGGGAATAC

NGGGAGGGCGGCGCCGAGCTCCTGCGCTCTCAGCGCACTCAGCACCGGGGCACGGA

[SEQ ID NO:12] Exon 2

TGAGCTCCACCGGGCCGGCGGCCGCTCTAGAACTAGTGGATCATGCCACTGTACCCTAGCCTGGGTAACAGAGTA

AGACACTATCTCTAAAAATAAAAAATAAGATAAAATATTTTTTTAAAAAAGAAACCATGTAATTTTCTCTTTTCTC

CCTACAG-GTATTG... .AGAAG-GTAATTAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTATGT

GTGTGCACAGCCTTATTAAGAATGTGATTGAGGTAAACATTATCTCCTATTCCCAAGGGGTAC

[SEO ID NO:13] Exon 3

AGATTTTTCCTTACTTTATTACCCCATCTGTATTTTGTAAAGTAGTATTTGAACCTAGTGTACACCTAACTTAG

TTGTATTCGTTGATGTTTACTTTGAATTATATAATGATTAGAAACATCTGACTTATCGTTCAATTTTTTCAG-

TTAA....CCAG-GTAAG

GATGAGCAGGGAAAAAAGTGGAGTTATGGTCATTAGGAAAAGATCCACTAGTTCTAGAGCGGCCGCCACGCCCGG

TGGAGCTT

[SEQ ID NO:14] Exon 4

CGCGGCGGCCGTTCTAGAACTAGTGGATCATACTCACGAAGACAGCAATTCTGTGTTCACTTTGGACATGTTGAA

TTTGAGACATAAAACACATTTTGCTGATGTTTGTTTCTAG-AACATA...GTCAAG-

GTCAGGTGCTCGTTGGGCTCTGCAGCAACCTGGT

ATCTTCCAACCTCTTAACGGGGCTNTACATAAGTGTTATCTTTTATATTAAGANTCATGGCTATTGGGCC

[SEQ ID NO:15] Exon 5

AATCTGTTACATACTTTGTTTGTGGAATCTGTCTTAATGTGTCTCACAAGCATCACAATTATTATTACTGTTAA

GTGTGTCATTTTATTTTCTTGAAAATATTTTAG-GT

GAGAA.....GCAGG-

GTAAGATTATATTCTGAGGTATTAATTTTTCCTTTTTT

AGAAGTACAGCATCATTTTTTCTTTCCAAATTAAGATGATAAAAATAATAAAATCACTGGTTTATTAAACATTA

CAGGTTGAGTATCCTTTATCCAAAATGTTTGGTATGAGAACTGTTTGGATTTTGGACTTTTTTGGATTTTGCAA

TATT

[SEO ID NO:16] Exon 6

CCGCAGCCCGGGGGATCACACAGACCTTAGTGGAAAATCTTCACTGGACCTGTGCCAAGAAGGGGGGTACATCTTC

24

p=0.01). These data are compatible with the possibility that the increased PC-1 content previously reported in skeletal muscle of insulin resistant subjects is, at least in part, due to reduction of PC-1 intracellular degradation, and its subsequent release into extracellular fluids, at the level of skeletal muscle tissue.

EXAMPLE 3

Intron/Exon Structure of PC-1

The nucleic acid sequences provided below are the intron-exon boundaries for the human PC-1 gene. It contains all the intron sequences immediately flanking the PC-1 exons. A few bases of the exon 5' and 3' are also provided, which are separated by a "-" sign from the intron, and are further in bold type.

The 3' flanking sequence to exon 2 (i.e. intron 3 at its 5' end) contains a GT repeat that is polymorphic, and provides a marker for genotyping of this locus. The sequences flanking the boundaries or crossing them are useful for specific amplification of the exons.

-continued

ATTGGATATGCTTGTCTTTGCTTCTTTAAACATTTTTTTTCTTTTCATTACCCAG-GTTTG...AAACTAA-
GTGAGTAACTTCAGAG

TTTACTGCTGGAATATCACCATTTCAGTGAGATTGACTAGGCAGGCAGTCTTTCTTGGAAAAGTACTGGCAGAAC
CTAACTGTTTCACTAAACTTTTCTAATGGGCAAAGTAGTTGAACCTTGTAAGGCGCCTTATCTTTAATAATGT
GA

[SEQ ID NO:17] Exon 7

TAAGAGAAAAATGAAGTCATCTTTAAGATIGGATTGTATCCACAGTGTGCTTTATAATTCATCCTGAATTTT
ATCTGATTAATAATCCCTCCTGGGTAATTTTTTTTACGTGATTAGACTGCTGTGGTACCCTGCTAAATGAGGTA
AGCCAATTGTCAGATGTATTTAATAACAATGTTTATTTTTTCCCTTCTAG-AAAAAATGT.....TCACC-
GTAAGCTCTGCATTTCAACTTCTATCTGTTTGAAGAAGTGAGATGGGATTGTAACATTTTTTGAGGGAATAGATT
TAAGATAAAAGAAAAACAACCTTATTTTCCAATAGGTAGTTAAGTAAGGAAACCCAGGTTCTGATCTTTGCTCTGC
CACAAACTAGCTGTGGCT

[SEQ ID NO:18] Exon 8

ACTACATAAAATCTTAAGAGTTGCGTTTGGCCATTACCTGATTTTTTGTGTTTTCTTTCCTTAACTTATTTATA
ATTCCATGTAGCTTCAGTTATCGGTTTCTTTTGATGATTTTTTTCGTGAATGTATTTAACATTAAGTAACAC
AACTTGCAATATAATCTGT
TTTATCTTTTTAG-GGATT...AACCA-GTGAGTCTTTGTTTTTCTACTAA
AATAGTTAATTATCTCATCTATTTCAATCAGAGTAAAATAACCAGATTCTCTAGAGCTTTTAATAACTGATTTC
ATTTAGTGTGCTGTGGCCAT

[SEQ ID NO:19] Exon 9

TAATCTCTGACTATTTAATATGTTGTTGCTGCTTAAGAGTCATATTACATGATTATTGTCGTCTAAGTGTGAAG
CTGTTGACCTTAAAGCATTCTAGCACTAGAGAGGAATGCATTGGTGTGGTATGAAAACATACTTTCCCTAAGAG
ATGAATGTTGCATGATTCTTCAATTTTCCCTTCATTTTCTGCTCCAG-ATTTGG.....AATGG-GTATGTG
AAATGAATTTTTCTAGGATCTGTAATATAGAACAGCTTATCTTATGTAATGTGTTTTTATTGAATCCTGAGC
TTTAGCATTTGAGTGATATGTTGGCTGAAAAATGAGAACTGAAGAACTCTTCTCAAAGAGTTTAGATAGATGGT
AATGGACAGTAAACTA

[SEQ ID NO:20] Exon 10

GGGAAATAAAGTTTTCAAATAAAACCCCTTGATTTCAACACAAATAGATGCGAAATAGCATTTACTAGCTCTTAA
TGACATTTTCAATGAAAAAACTATATTTTACACCCAAACAATGTGAGCCATCTTTATTTTTGTTGTTCTTC
ATTTTAG-TTCAGTA...AGATGAAAG-GTCTGTAGGCAATTAATTTCTATGTAATAACTTCGTTTTGTA
GAAATGATATACTATTTTCCCTAGACTACAACAAACTTTGCTATTGCTATGATGTTTTATATCGAAATAAAT
TCTTTAGTAAATGATC

[SEQ ID NO:21] Exon 11

GAATTTCAAAGCTGTAATTAATTTCTCAGTAGAACTGTTACACCAGTGTATATAAATTAATCCCTATCAATTGA
GGAATTATTTTTCCATTCTGTTTTCAATGTGTTGTAATAATATTACATTTTGATACTGTTTGATTAG-
ACCACA...CAGTGA[1]-GTAA
GTACATTTTTCTCAGTAATTATTTTATTAAACCCAGTCATCGGGCTGAACCTCGCTTTGAAGGAGGCTGCTAGAC
CATTTTATAAGATTCATCATTTCTGAAAAAGCAAGTATTATACACAATATTACTAAATATAAGGATGCACTTT
AAACAAATAAGAGTTGG

[SEQ ID NO:22] Exon 12

GTCTTAGTTTAATGTGAATCAGCTCATTTGTAGTTGCATCCACTGGCCCAATCTATCAATCTGTGGTCTTTCTT
TCTTCTTTGTTTCTTTCTTTTTTTTTTTTTTAAACAGAGATAGCTTTATGTATAAATAGCCATTAGTGTGAAG
GTATCACATGAGGTTGTGCTTCCCATCTTAG-GTCATC.....ATCATG-GTAATCTGAATTTGCATTA

-continued

TTTACTCTTCAGGATAAAGGGCTGAAGAAAGTTTACTTGATGGTTTCCCAATTTTTGTGAATGTTGTAGTTAAT
TCTTTTTTAAAAATGTAGTTTCTTATGGACAGTCTTTAGGAAAAAATACATTAAATATAAAATATAAGTGAAAC
ACAGAATTCACAGAAACC

[SEQ ID NO:23] Exon 13

GATTTTGAAGAAAGTGAAGTGATAGGTACAGCTGAAATTCGTCTTACCTATCAGATCTTCAACTAATATGAGTG
CTACACCCATGTTTAAACGAATTTAACCTTGGAAGTGAAAGAAGTTCTGCTCTGCATATTAATTTTTTGTAAAG
TTACAGCATGTTTGGGATTTTTTTTTCTCCTAG-GCATGG....TACTATTCAT-GTAAGTATATCTC
TGTGATAACTTTGAATATGGTCATATTAAGAATACCTTCTTTAGGCCGGGCACAGTGGCTCATGCCTGTAATCG
CAGCACTTTGGGAGGCCAAAGTGGGTGGTCACCTGAGGTCAGGAGTTCGAGACCAGCCTGGCCACATGGTGAAA
CCCTAAAAATACATATAC

[SEQ ID NO:24] Exon 14

GATCCAACTCTGCATTTAAATACCAAGGCAGGTTTAAAGAGTTCATTTAAGTCATTACATTGTAGCCACTGAA
AGGAATTAGACAGACCTTTAGGGATCTGACATCTTATTTTTTGTATTATGTTTAAATATAGTATACAATCAAAC
TATTAATTCCTTATGTTTGTCCCTCCAG-TTAACTATGAAGGCATTGCCCGAAATCTTTCT-GTGAGTA
TCTTTATTTTCCATTATCTAGTTATTTTACTTTTGTATAATATATATGAGAGAAAAGTTTCAGCATCTATTAT
TGGGATGAAGGATTAGAATATTTTAGTAATCTGGGCCAACATGGAATGCTGTGTAGTTTAAAGATC

[SEQ ID NO:25] Exon 15

CTGATGAAATGTTTGTGAAAAAAATTTTCATATGAAGTTAGAAAGCAATTTCAAGAAAAGTTGACACTTTTATA
GATATTAGGAAATATCTTTCCCTAATAAATATCTTTCCCTAAAAAGTTGACACTTTTTTAGATATTAGGGAAA
TAATAGTTTTCTTTGCT
GTTTGCATTTTCAG-TGCCGGG....GCATT-GTAAGTCTGACAGTCTCCAG
GTAAACTTAGTCTGATCGGTTAGTGATTACAGGTAACCATTTGGGCCCTTTCTAACAATATTGTTATGTGAAAAC
GTATAAGTATGATTCCTTCACTCTAACCCAGGATTTCTAATGTCGGCCTATGGATGTTTGAGTTAGATAATTCT
TTGTTGTGGAGAGCTGTC

[SEQ ID NO:26] Exon 16

AAAAGATAGAGGTGACTTCTTAATGCTTTTCAAAGCCAGGTGGTTTTATTACCCTTGTGTTGGTTTAAACAAAT
AGTTACATACTTTTAAATCAATGAAAATAATGTTATG
ATTATCAATTATGTTTATGAAAGGACTTTACATTTTAAATTCATATATGTCAACATTAG-GAAT....GCAA-
TCTAAGAAAAAATGATATGCAAAGTTTACACTTGAAACATACTGTGATTATATGTCTTGAATGAGAATTAAT
GGAACATACTTTTCAAAAGCTATTTTCTTTGAACATTAAAGAATTTTGTAAAGTTTATATTTCATTGGCTATT
ACTAAAAAGTCAAAAAAC

[SEQ ID NO:27] Exon 17

AAAACTAAGAGACCTATCCTAGATGTCCTTAGATTATGTGTGTGATAGGGTTAAACATATTTCCACAAAGTC
CACTGAGCGTGGTAGTTTTCCTCTTATCTTATCATAACAGTTTGTATATGTACAATGTGGATAACAGAATTTT
GGGACCAACTGTAGACAGCTGAAATGCACTGATAAACTTCTTTTCTGGCCATCTAG-GCCCT....GTG
TG-GTAAGTGTGAACAGGTGCCTTTTTCCTTCTGAAAATAGACCTGAAATAGGA
TTATCAAAAGCAGGTACATTGTAGGCACTTTGTGGAGATGATGGTGAGGCAAGACAGATTTTACCTTCTTCC
TGACTCTCAGACTCACTGAAGAAATGTGGGAACATG

[SEQ ID NO:28] Exon 18

CATATCAGTATTTCTATTTAAAAATAACCTAGTCTTAAATACTCTAAACCCAAGAGAGTTTATACTTTTATTTT
AGTTAAAGAGTAAATGACTCATGTATTTGGTTTTAAAAAAGTAAAGATCATGGCACAAGTCTACTATTTGTTTGA
TTTGAACATCTAAGTAACCTACCATCTTGAAATTATGCAG-ATTTA....CTTCG-GTAAGTATCGTCAA
GAAGTTTGGTCCAGTATGTATGTTTGATAGCACCTCTGCATAGCATGTGCTGAAAAATACTTAAATAATCAAA

-continued

TTAgAATTTAGGAGTGGGGTAGGTAAACATATGTTTAAATCTAGGGGCGCATGTAAATCTTTGTGATATAT
CTTTCTCTTTCTAGTTT

[SEQ ID NO:29] Exon 19

GTGAAAGAGCAACACTCTTGCCCTTGAAAGAGAAAAAATCCACTAATACAAGACTATCATAAATGATCTTTGT
TTTATGTTGGAATAATCAATCTATAGCGGTCTATGTTACAAAATTTAAACATGTCCTCAGTCCTTACAAATAG
TTTTATAACCTTTTTCAG-ATTTTGGCC...GAAG-GTAAGGCATGCTACACACTCAAGCTCGGAATGTG
AAGCAGGCATTTTCTCATCAGTGTGAAATGCAGAGAACTGGCTTGGGGGTATTATTGAGAATAACCAATAAAT
AAAGGGAGTTCTGGAGGACCACCTGATGAAACATAGAGGTTTCTTTGCT

[SEQ ID NO:30] Exon 20

GTCTTCTTAATTTGTTATGCTTGTAATCAGTTTTTTAATAGTTAAAAGTAAATCTTCAATATAAT
TAAGTAGAGGAAAGGATTAGATGAGTGTATCACACTATATATTATCATATAATGCACACTAACTACATTATT
CATCTGTGACCCAAG-A GAAGATTA...GACAGAAAT-GCAAGTATTTGTCACCTCTTTATGTGTGGCC
ATTTCAAATTAATGATTAAGCAGAACATTAAATGCATAGTTTCTCACTGTTACCTTGGCTTTTACTCAGTTCC
CGCATTAGAGGAACACTGAAGAGGGAGTCAGAAAAAT

[SEQ ID NO:31] Exon 21

TTTAATATTGTAAAGCATTTTACACTTTAGTTAGAAAAAAGATGAATATACTAGTAGGAAAATAGGGAAGGAC
ATGAGCTGACAGCTAGAGCTTCATAATTTTATGATGTAGTTCACCTTTAAATATTAATAAAGCAATTTTCTTCTC
TGTGCCGTATATCTGAGAGTTCTTCTCATTTTCGTTCTTCAG-GACA...CCACCAC-GTAAGTTTTC
TCTCTGACCTTCCCTTTTCTCCTTTTGTCTTCTTGTGTTTATAAATCCTACCATACATTATAGGTAATAT
ATATATTACCTATTATATATATATAGCTATATATATATACCTTGTATTATTATTGTGA

[SEQ ID NO:32] Exon 22

CTCATCTTGAAAGACTTCTTAAATATTTATTTTGTAAAGGACTTGACCAACACATAACATTTTCCCTCGAC
CCTGTACTTGGGAAAGTTTACAGGTTTAAGATGGTACTCAGCTAATTTTAAAAATGCTCCCTAACCATGAGA
AAGTATAATTTCTATGTTATTGTGAGAATGAAAAAGTTGTCTCTTTTCTCTTTGTAG-AACTA...TT
CAAG-GTAAATAATGTTAACTCTATATTTGATAATTTTAATGAATTTGTGCACAT
ATAGGCATAATTCATATGTATAGGACTTATGGTCTAAATTAATGAATTAATACCAATACATTCTTAAAGGTTT
AACTTTGAGAATACTAGTACACAAAAATCTAC

[SEQ ID NO:33] Exon 23

CTGGTGATATAGCACGACTCTGTCTTAAACAAAAACAAAACAAACGAAGACTGAAGCCAACTTGACTTTA
TCTTTATTTACTATAAATGCTAATTTTGAATCATGGTGTTAATTTATTTACACGTCACATGGTCCCTTGTCT
TTTGAAACTACACTGGCTTCTATCTGTTCAG-TTATA...GAGGCA-GTAAGAACATATTTCACTACTC
TTAAAAATAGGAATTACCATCCAGTAGAAATGGGATTACCATCCAGTTGAGTCAAGAGAACCTTTTATCCAGT
GTCGTATGTTTATGTGTATGACACTTCTGACTACACAGGAAGCCTCTTGAATATCTGATTAAATTTGATGTtTT
GCTCAATGTTCAGTAAAA

[SEQ ID NO:34] Exon 24

GTTCCTTATATTTAATTTATGGTTGGAATTTGATTTTATATGTATTTAAAGCATGCTCTACTGAAATATTCATCA
AAAGGAAGATAGTTATTTCTTTTAAATGAATATTGGCATGTTTACAG-AAAAA...TGTGTG-
GTAAGTAGCTTTGTATATTTAC
TTTGCATGTTGAAATCTAGACATATGCATATTGTTTATGTACCCCATCTGACATTACAGTGAGAGAAAGCACA
ACTGAGTACACATGGACTTCGAAATTTAGGATGCTTTTAAATTTGATCTTTTAAGATGACATATCTTTGGGGAA
GACTACCCTGTCTGCTTT

[SEQ ID NO:35] Exon 25

AATTAAACAACATGCATGGTATGTATTAGAAGGAAAGCTACTCAAGAGGAGAGATGATGCCATAACAAATCATGT

-continued

GGCACGTTCCACTTCAGAGCTGAAATCTCGTAAATGATTAAACTGGGGAGATGGAGCACTTATAGAAGTGAAC TG
 AGTGTCTCTTGCTAACCTTTCTTTTATATTTCTTCTCTAG-CATGG...ATTAA-AAAAGAAAA
 TATTCTATCTCTGCTCACTGGTAATTAACATAGGTTTAAATGGCTTCAAATGTGGCCCTATAGACGGTTAAAAAT
 TGTACCTTATCTTGCAAACTTCAGAGCACCAGTCAGTGCATGCAAGGTGCCATTTTATTGAGATGCTTAGA
 ATGTTCTCTTCTGTGCAC

It is to be understood that this invention is not limited to the particular methodology, protocols, formulations and reagents described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a", "and", and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a complex" includes a plurality of such complexes and reference to "the formulation" includes reference to one or more formulations and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly

understood to one of ordinary skill in the art to which this invention belongs. Although any methods, devices and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods, devices and materials are now described.

All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing, for example, the cell lines, constructs, and methodologies that are described in the publications which might be used in connection with the presently described invention. The publications discussed above and throughout the text are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 35

<210> SEQ ID NO 1
 <211> LENGTH: 3486
 <212> TYPE: DNA
 <213> ORGANISM: H. sapiens
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (164)...(2785)

<400> SEQUENCE: 1

ggccacgatg gagegcgacg gctgcgcggg gggcgggagc cgcggcgggc agggcggggc	60
cgtccccgg gagggcccg cggggaacgg ccgcgatcgg ggccgcagcc acgctgccga	120
ggcggccggg gaccgcgagg cggccgcgtc cttgctggcc cct atg gac gtg ggg	175
Met Asp Val Gly	
1	
gag gag ccg ctg gag aag gcg gcg cgc gcc cgc act gcc aag gac ccc	223
Glu Glu Pro Leu Glu Lys Ala Ala Arg Ala Arg Thr Ala Lys Asp Pro	
5 10 15 20	
aac acc tat aaa gta ctc tcg ctg gta ttg tca gta tgt gtg tta aca	271
Asn Thr Tyr Lys Val Leu Ser Leu Val Leu Ser Val Cys Val Leu Thr	
25 30 35	
aca ata ctt ggt tgt ata ttt ggg ttg aaa cca agc tgt gcc aaa gaa	319
Thr Ile Leu Gly Cys Ile Phe Gly Leu Lys Pro Ser Cys Ala Lys Glu	
40 45 50	
gtt aaa agt tgc aaa ggt cgc tgt ttc gag aga aca ttt ggg aac tgt	367
Val Lys Ser Cys Lys Gly Arg Cys Phe Glu Arg Thr Phe Gly Asn Cys	
55 60 65	
cgc tgt gat gct gcc tgt gtt gag ctt gga aac tgc tgt tta gat tac	415
Arg Cys Asp Ala Ala Cys Val Glu Leu Gly Asn Cys Cys Leu Asp Tyr	
70 75 80	

-continued

cag gag acg tgc ata gaa cca gaa cat ata tgg act tgc aac aaa ttc Gln Glu Thr Cys Ile Glu Pro Glu His Ile Trp Thr Cys Asn Lys Phe 85 90 95 100	463
agg tgt ggt gag aaa agg ttg acc aga agc ctc tgt gcc tgt tca gat Arg Cys Gly Glu Lys Arg Leu Thr Arg Ser Leu Cys Ala Cys Ser Asp 105 110 115	511
gac tgc aag gac aag ggc gac tgc tgc atc aac tac agt tct gtg tgt Asp Cys Lys Asp Lys Gly Asp Cys Cys Ile Asn Tyr Ser Ser Val Cys 120 125 130	559
caa ggt gag aaa agt tgg gta gaa gaa cca tgt gag agc att aat gag Gln Gly Glu Lys Ser Trp Val Glu Glu Pro Cys Glu Ser Ile Asn Glu 135 140 145	607
cca cag tgc cca gca ggg ttt gaa acg cct cct acc ctc tta ttt tct Pro Gln Cys Pro Ala Gly Phe Glu Thr Pro Pro Thr Leu Leu Phe Ser 150 155 160	655
ttg gat gga ttc agg gca gaa tat tta cac act tgg ggt gga ctt ctt Leu Asp Gly Phe Arg Ala Glu Tyr Leu His Thr Trp Gly Gly Leu Leu 165 170 175 180	703
cct gtt att agc aaa cta aaa aaa tgt gga aca tat act aaa aac atg Pro Val Ile Ser Lys Leu Lys Lys Cys Gly Thr Tyr Thr Lys Asn Met 185 190 195	751
aga ccg gta tat cca aca aaa act ttc ccc aat cac tac agc att gtc Arg Pro Val Tyr Pro Thr Lys Thr Phe Pro Asn His Tyr Ser Ile Val 200 205 210	799
acc gga ttg tat cca gaa tct cat ggc ata atc gac aat aaa atg tat Thr Gly Leu Tyr Pro Glu Ser His Gly Ile Ile Asp Asn Lys Met Tyr 215 220 225	847
gat ccc aaa atg aat gct tcc ttt tca ctt aaa agt aaa gag aaa ttt Asp Pro Lys Met Asn Ala Ser Phe Ser Leu Lys Ser Lys Glu Lys Phe 230 235 240	895
aat cct gag tgg tac aaa gga gaa cca att tgg gtc aca gct aag tat Asn Pro Glu Trp Tyr Lys Gly Glu Pro Ile Trp Val Thr Ala Lys Tyr 245 250 255 260	943
caa ggc ctc aag tct ggc aca ttt ttc tgg cca gga tca gat gtg gaa Gln Gly Leu Lys Ser Gly Thr Phe Phe Trp Pro Gly Ser Asp Val Glu 265 270 275	991
att aac gga att ttc cca gac atc tat aaa atg tat aat ggt tca gta Ile Asn Gly Ile Phe Pro Asp Ile Tyr Lys Met Tyr Asn Gly Ser Val 280 285 290	1039
cca ttt gaa gaa agg att tta gct gtt ctt cag tgg cta cag ctt cct Pro Phe Glu Glu Arg Ile Leu Ala Val Leu Gln Trp Leu Gln Leu Pro 295 300 305	1087
aaa gat gaa aga cca cac ttt tac act ctg tat tta gaa gaa cca gat Lys Asp Glu Arg Pro His Phe Tyr Thr Leu Tyr Leu Glu Glu Pro Asp 310 315 320	1135
tct tca ggt cat tca tat gga cca gtc agc agt gaa gtc atc aaa gcc Ser Ser Gly His Ser Tyr Gly Pro Val Ser Ser Glu Val Ile Lys Ala 325 330 335 340	1183
ttg cag agg gtt gat ggt atg gtt ggt atg ctg atg gat ggt ctg aaa Leu Gln Arg Val Asp Gly Met Val Gly Met Leu Met Asp Gly Leu Lys 345 350 355	1231
gag ctg aac ttg cac aga tgc ctg aac ctc atc ctt att tca gat cat Glu Leu Asn Leu His Arg Cys Leu Asn Leu Ile Leu Ile Ser Asp His 360 365 370	1279
ggc atg gaa caa ggc agt tgt aag aaa tac ata tat ctg aat aaa tat Gly Met Glu Gln Gly Ser Cys Lys Lys Tyr Ile Tyr Leu Asn Lys Tyr 375 380 385	1327
ttg ggg gat gtt aaa aat att aaa gtt atc tat gga cct gca gct cga Leu Gly Asp Val Lys Asn Ile Lys Val Ile Tyr Gly Pro Ala Ala Arg 390 395 400	1375

-continued

ttg aga ccc tct gat gtc cca gat aaa tac tat tca ttt aac tat gaa Leu Arg Pro Ser Asp Val Pro Asp Lys Tyr Tyr Ser Phe Asn Tyr Glu 405 410 415 420	1423
ggc att gcc cga aat ctt tct tgc cgg gaa cca aac cag cac ttc aaa Gly Ile Ala Arg Asn Leu Ser Cys Arg Glu Pro Asn Gln His Phe Lys 425 430 435	1471
cct tac ctg aaa cat ttc tta cct aag cgt ttg cac ttt gct aag agt Pro Tyr Leu Lys His Phe Leu Pro Lys Arg Leu His Phe Ala Lys Ser 440 445 450	1519
gat aga att gag ccc ttg aca ttc tat ttg gac cct cag tgg caa ctt Asp Arg Ile Glu Pro Leu Thr Phe Tyr Leu Asp Pro Gln Trp Gln Leu 455 460 465	1567
gca ttg aat ccc tca gaa agg aaa tat tgt gga agt gga ttt cat ggc Ala Leu Asn Pro Ser Glu Arg Lys Tyr Cys Gly Ser Gly Phe His Gly 470 475 480	1615
tct gac aat gta ttt tca aat atg caa gcc ctc ttt gtt ggc tat gga Ser Asp Asn Val Phe Ser Asn Met Gln Ala Leu Phe Val Gly Tyr Gly 485 490 495 500	1663
cct gga ttc aag cat ggc att gag gct gac acc ttt gaa aac att gaa Pro Gly Phe Lys His Gly Ile Glu Ala Asp Thr Phe Glu Asn Ile Glu 505 510 515	1711
gtc tat aac tta atg tgt gat tta ctg aat ttg aca ccg gct cct aat Val Tyr Asn Leu Met Cys Asp Leu Leu Asn Leu Thr Pro Ala Pro Asn 520 525 530	1759
aac gga act cat gga agt ctt aac cac ctt cta aag aat cct gtt tat Asn Gly Thr His Gly Ser Leu Asn His Leu Leu Lys Asn Pro Val Tyr 535 540 545	1807
acg cca aag cat ccc aaa gaa gtg cac ccc ctg gta cag tgc ccc ttc Thr Pro Lys His Pro Lys Glu Val His Pro Leu Val Gln Cys Pro Phe 550 555 560	1855
aca aga aac ccc aga gat aac ctt ggc tgc tca tgt aac cct tcg att Thr Arg Asn Pro Arg Asp Asn Leu Gly Cys Ser Cys Asn Pro Ser Ile 565 570 575 580	1903
ttg ccg att gag gat ttt caa aca cag ttc aat ctg act gtg gca gaa Leu Pro Ile Glu Asp Phe Gln Thr Gln Phe Asn Leu Thr Val Ala Glu 585 590 595	1951
gag aag att att aag cat gaa act tta ccc tat gga aga cct aga gtt Glu Lys Ile Ile Lys His Glu Thr Leu Pro Tyr Gly Arg Pro Arg Val 600 605 610	1999
ctc cag aag gaa aac acc atc tgt ctt ctt tcc cag cac cag ttt atg Leu Gln Lys Glu Asn Thr Ile Cys Leu Leu Ser Gln His Gln Phe Met 615 620 625	2047
agt gga tac agc caa gac atc tta atg ccc ctt tgg aca tcc tat acc Ser Gly Tyr Ser Gln Asp Ile Leu Met Pro Leu Trp Thr Ser Tyr Thr 630 635 640	2095
gtg gac aga aat gac agt ttc tct acg gaa gac ttc tcc aac tgt ctg Val Asp Arg Asn Asp Ser Phe Ser Thr Glu Asp Phe Ser Asn Cys Leu 645 650 655 660	2143
tac cag gac ttt aga att cct ctt agt cct gtc cat aaa tgt tca ttt Tyr Gln Asp Phe Arg Ile Pro Leu Ser Pro Val His Lys Cys Ser Phe 665 670 675	2191
tat aaa aat aac acc aaa gtg agt tac ggg ttc ctc tcc cca cca caa Tyr Lys Asn Asn Thr Lys Val Ser Tyr Gly Phe Leu Ser Pro Pro Gln 680 685 690	2239
cta aat aaa aat tca agt gga ata tat tct gaa gct ttg ctt act aca Leu Asn Lys Asn Ser Ser Gly Ile Tyr Ser Glu Ala Leu Leu Thr Thr 695 700 705	2287
aat ata gtg cca atg tac cag agt ttt caa gtt ata tgg cgc tac ttt Asn Ile Val Pro Met Tyr Gln Ser Phe Gln Val Ile Trp Arg Tyr Phe	2335

-continued

710	715	720	
cat gac acc cta ctg cga aag tat gct gaa gaa aga aat ggt gtc aat			2383
His Asp Thr Leu Leu Arg Lys Tyr Ala Glu Glu Arg Asn Gly Val Asn			
725	730	735	740
gtc gtc agt ggt cct gtg ttt gac ttt gat tat gat gga cgt tgt gat			2431
Val Val Ser Gly Pro Val Phe Asp Phe Asp Tyr Asp Gly Arg Cys Asp			
745	750	755	
tcc tta gag aat ctg agg caa aaa aga aga gtc atc cgt aac caa gaa			2479
Ser Leu Glu Asn Leu Arg Gln Lys Arg Arg Val Ile Arg Asn Gln Glu			
760	765	770	
att ttg att cca act cac ttc ttt att gtg cta aca agc tgt aaa gat			2527
Ile Leu Ile Pro Thr His Phe Phe Ile Val Leu Thr Ser Cys Lys Asp			
775	780	785	
aca tct cag acg cct ttg cac tgt gaa aac cta gac acc tta gct ttc			2575
Thr Ser Gln Thr Pro Leu His Cys Glu Asn Leu Asp Thr Leu Ala Phe			
790	795	800	
att ttg cct cac agg act gat aac agc gag agc tgt gtg cat ggg aag			2623
Ile Leu Pro His Arg Thr Asp Asn Ser Glu Ser Cys Val His Gly Lys			
805	810	815	820
cat gac tcc tca tgg gtt gaa gaa ttg tta atg tta cac aga gca cgg			2671
His Asp Ser Ser Trp Val Glu Glu Leu Leu Met Leu His Arg Ala Arg			
825	830	835	
atc aca gat gtt gag cac atc act gga ctc agc ttc tat caa caa aga			2719
Ile Thr Asp Val Glu His Ile Thr Gly Leu Ser Phe Tyr Gln Gln Arg			
840	845	850	
aaa gag cca gtt tca gac att tta aag ttg aaa aca cat ttg cca acc			2767
Lys Glu Pro Val Ser Asp Ile Leu Lys Leu Lys Thr His Leu Pro Thr			
855	860	865	
ttt agc caa gaa gac tga tatgtttttt atccccaaac accatgaatc			2815
Phe Ser Gln Glu Asp			
870			
tttttgagag aaccttatat tttatatagt cctctagcta cactattgca ttgttcagaa			2875
actgtcgacc agagttagaa cggagccctc ggtgatgcgg acatctcagg gaaacttcg			2935
tactcagcac agcagtgagg agtgttcctg ttgaatcttg cacatatattg aatgtgtaag			2995
cattgtatac attgatcaag ttcgggggaa taaagacaga ccacacctaa aactgccttt			3055
ctgcttctct taaaggagaa gtactgtgtg acattgtctg gataccagat atttgaatct			3115
ttcttactat tggtaataaa ccttgatggc attggggcaa cagtagactt atagtaggg			3175
tggggtagcc catgttatgt gactatcttt atgagaattt taaagtgggt ctggatatct			3235
tttaacttgg agtttcattt cttttcattg taatcaaaaa aaaaattaac agaagccaaa			3295
atacttctga gaccttgtt caatctttgc tgtatatccc ctcaaatcc aagttattaa			3355
tcttatgtgt tttcttttta attttttgat tggatttctt tagatttaat ggttcaaatg			3415
agttcaactt tgagggacga tctttgaata tacttaccta ttataaaatc ttactttgta			3475
tttgatttta a			3486

<210> SEQ ID NO 2
 <211> LENGTH: 873
 <212> TYPE: PRT
 <213> ORGANISM: H. sapiens

<400> SEQUENCE: 2

Met Asp Val Gly Glu Glu Pro Leu Glu Lys Ala Ala Arg Ala Arg Thr
 1 5 10 15

Ala Lys Asp Pro Asn Thr Tyr Lys Val Leu Ser Leu Val Leu Ser Val
 20 25 30

-continued

Cys Val Leu Thr Thr Ile Leu Gly Cys Ile Phe Gly Leu Lys Pro Ser
 35 40 45
 Cys Ala Lys Glu Val Lys Ser Cys Lys Gly Arg Cys Phe Glu Arg Thr
 50 55 60
 Phe Gly Asn Cys Arg Cys Asp Ala Ala Cys Val Glu Leu Gly Asn Cys
 65 70 75 80
 Cys Leu Asp Tyr Gln Glu Thr Cys Ile Glu Pro Glu His Ile Trp Thr
 85 90 95
 Cys Asn Lys Phe Arg Cys Gly Glu Lys Arg Leu Thr Arg Ser Leu Cys
 100 105 110
 Ala Cys Ser Asp Asp Cys Lys Asp Lys Gly Asp Cys Cys Ile Asn Tyr
 115 120 125
 Ser Ser Val Cys Gln Gly Glu Lys Ser Trp Val Glu Glu Pro Cys Glu
 130 135 140
 Ser Ile Asn Glu Pro Gln Cys Pro Ala Gly Phe Glu Thr Pro Pro Thr
 145 150 155 160
 Leu Leu Phe Ser Leu Asp Gly Phe Arg Ala Glu Tyr Leu His Thr Trp
 165 170 175
 Gly Gly Leu Leu Pro Val Ile Ser Lys Leu Lys Lys Cys Gly Thr Tyr
 180 185 190
 Thr Lys Asn Met Arg Pro Val Tyr Pro Thr Lys Thr Phe Pro Asn His
 195 200 205
 Tyr Ser Ile Val Thr Gly Leu Tyr Pro Glu Ser His Gly Ile Ile Asp
 210 215 220
 Asn Lys Met Tyr Asp Pro Lys Met Asn Ala Ser Phe Ser Leu Lys Ser
 225 230 235 240
 Lys Glu Lys Phe Asn Pro Glu Trp Tyr Lys Gly Glu Pro Ile Trp Val
 245 250 255
 Thr Ala Lys Tyr Gln Gly Leu Lys Ser Gly Thr Phe Phe Trp Pro Gly
 260 265 270
 Ser Asp Val Glu Ile Asn Gly Ile Phe Pro Asp Ile Tyr Lys Met Tyr
 275 280 285
 Asn Gly Ser Val Pro Phe Glu Glu Arg Ile Leu Ala Val Leu Gln Trp
 290 295 300
 Leu Gln Leu Pro Lys Asp Glu Arg Pro His Phe Tyr Thr Leu Tyr Leu
 305 310 315 320
 Glu Glu Pro Asp Ser Ser Gly His Ser Tyr Gly Pro Val Ser Ser Glu
 325 330 335
 Val Ile Lys Ala Leu Gln Arg Val Asp Gly Met Val Gly Met Leu Met
 340 345 350
 Asp Gly Leu Lys Glu Leu Asn Leu His Arg Cys Leu Asn Leu Ile Leu
 355 360 365
 Ile Ser Asp His Gly Met Glu Gln Gly Ser Cys Lys Lys Tyr Ile Tyr
 370 375 380
 Leu Asn Lys Tyr Leu Gly Asp Val Lys Asn Ile Lys Val Ile Tyr Gly
 385 390 395 400
 Pro Ala Ala Arg Leu Arg Pro Ser Asp Val Pro Asp Lys Tyr Tyr Ser
 405 410 415
 Phe Asn Tyr Glu Gly Ile Ala Arg Asn Leu Ser Cys Arg Glu Pro Asn
 420 425 430
 Gln His Phe Lys Pro Tyr Leu Lys His Phe Leu Pro Lys Arg Leu His
 435 440 445

Phe 450	Ala	Lys	Ser	Asp	Arg	Ile 455	Glu	Pro	Leu	Thr	Phe 460	Tyr	Leu	Asp	Pro
Gln 465	Trp	Gln	Leu	Ala	Leu 470	Asn	Pro	Ser	Glu	Arg 475	Lys	Tyr	Cys	Gly	Ser 480
Gly	Phe	His	Gly	Ser 485	Asp	Asn	Val	Phe	Ser 490	Asn	Met	Gln	Ala	Leu 495	Phe
Val	Gly	Tyr	Gly 500	Pro	Gly	Phe	Lys	His 505	Gly	Ile	Glu	Ala	Asp 510	Thr	Phe
Glu	Asn	Ile 515	Glu	Val	Tyr	Asn	Leu 520	Met	Cys	Asp	Leu 525	Leu	Asn	Leu	Thr
Pro 530	Ala	Pro	Asn	Asn	Gly	Thr 535	His	Gly	Ser	Leu	Asn 540	His	Leu	Leu	Lys
Asn 545	Pro	Val	Tyr	Thr	Pro 550	Lys	His	Pro	Lys	Glu 555	Val	His	Pro	Leu	Val 560
Gln	Cys	Pro	Phe	Thr 565	Arg	Asn	Pro	Arg	Asp 570	Asn	Leu	Gly	Cys	Ser 575	Cys
Asn	Pro	Ser	Ile 580	Leu	Pro	Ile	Glu	Asp 585	Phe	Gln	Thr	Gln	Phe 590	Asn	Leu
Thr	Val	Ala 595	Glu	Glu	Lys	Ile	Ile 600	Lys	His	Glu	Thr 605	Leu	Pro	Tyr	Gly
Arg 610	Pro	Arg	Val	Leu	Gln	Lys 615	Glu	Asn	Thr	Ile 620	Cys	Leu	Leu	Ser	Gln
His 625	Gln	Phe	Met	Ser	Gly 630	Tyr	Ser	Gln	Asp	Ile 635	Leu	Met	Pro	Leu	Trp 640
Thr	Ser	Tyr	Thr 645	Val	Asp	Arg	Asn	Asp	Ser 650	Phe	Ser	Thr	Glu	Asp 655	Phe
Ser	Asn	Cys	Leu 660	Tyr	Gln	Asp	Phe	Arg 665	Ile	Pro	Leu	Ser	Pro 670	Val	His
Lys 675	Cys	Ser	Phe	Tyr	Lys	Asn	Asn 680	Thr	Lys	Val	Ser 685	Tyr	Gly	Phe	Leu
Ser 690	Pro	Pro	Gln	Leu	Asn	Lys 695	Asn	Ser	Ser	Gly 700	Ile	Tyr	Ser	Glu	Ala
Leu 705	Leu	Thr	Thr	Asn	Ile 710	Val	Pro	Met	Tyr	Gln 715	Ser	Phe	Gln	Val	Ile 720
Trp	Arg	Tyr	Phe 725	His	Asp	Thr	Leu	Leu	Arg 730	Lys	Tyr	Ala	Glu	Glu 735	Arg
Asn	Gly	Val	Asn 740	Val	Val	Ser	Gly	Pro 745	Val	Phe	Asp	Phe	Asp 750	Tyr	Asp
Gly	Arg	Cys 755	Asp	Ser	Leu	Glu	Asn 760	Leu	Arg	Gln	Lys 765	Arg	Arg	Val	Ile
Arg 770	Asn	Gln	Glu	Ile	Leu	Ile 775	Pro	Thr	His	Phe 780	Phe	Ile	Val	Leu	Thr
Ser 785	Cys	Lys	Asp	Thr	Ser 790	Gln	Thr	Pro	Leu	His 795	Cys	Glu	Asn	Leu	Asp 800
Thr	Leu	Ala	Phe 805	Ile	Leu	Pro	His	Arg 810	Thr	Asp	Asn	Ser	Glu	Ser 815	Cys
Val	His	Gly	Lys 820	His	Asp	Ser	Ser	Trp 825	Val	Glu	Glu	Leu	Leu	Met	Leu
His	Arg	Ala 835	Arg	Ile	Thr	Asp	Val 840	Glu	His	Ile	Thr 845	Gly	Leu	Ser	Phe
Tyr 850	Gln	Gln	Arg	Lys	Glu	Pro 855	Val	Ser	Asp	Ile 860	Leu	Lys	Leu	Lys	Thr
His	Leu	Pro	Thr	Phe	Ser	Gln	Glu	Asp							

-continued

865 870

<210> SEQ ID NO 3
 <211> LENGTH: 3486
 <212> TYPE: DNA
 <213> ORGANISM: H. sapiens
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (164)...(2785)

<400> SEQUENCE: 3

ggccacgatg gagcgcgacg gctgcgcggg gggcgggagc cgcggcgcg agggcgggcg 60
 cgctccccgg gagggcccg cggggaacgg ccgcgacgg ggcgcagcc acgctgccga 120
 ggcgccccgg gaccgcagg cgccgcgtc ctgctggcc cct atg gac gtg ggg 175
 Met Asp Val Gly
 1

gag gag ccg ctg gag aag gcg gcg cgc gcc cgc act gcc aag gac ccc 223
 Glu Glu Pro Leu Glu Lys Ala Ala Arg Ala Arg Thr Ala Lys Asp Pro
 5 10 15 20

aac acc tat aaa gta ctc tcg ctg gta ttg tca gta tgt gtg tta aca 271
 Asn Thr Tyr Lys Val Leu Ser Leu Val Leu Ser Val Cys Val Leu Thr
 25 30 35

aca ata ctt ggt tgt ata ttt ggg ttg aaa cca agc tgt gcc aaa gaa 319
 Thr Ile Leu Gly Cys Ile Phe Gly Leu Lys Pro Ser Cys Ala Lys Glu
 40 45 50

gtt aaa agt tgc aaa ggt cgc tgt ttc gag aga aca ttt ggg aac tgt 367
 Val Lys Ser Cys Lys Gly Arg Cys Phe Glu Arg Thr Phe Gly Asn Cys
 55 60 65

cgc tgt gat gct gcc tgt gtt gag ctt gga aac tgc tgt tta gat tac 415
 Arg Cys Asp Ala Ala Cys Val Glu Leu Gly Asn Cys Cys Leu Asp Tyr
 70 75 80

cag gag acg tgc ata gaa cca gaa cat ata tgg act tgc aac aaa ttc 463
 Gln Glu Thr Cys Ile Glu Pro Glu His Ile Trp Thr Cys Asn Lys Phe
 85 90 95 100

agg tgt ggt gag aaa agg ttg acc aga agc ctc tgt gcc tgt tca gat 511
 Arg Cys Gly Glu Lys Arg Leu Thr Arg Ser Leu Cys Ala Cys Ser Asp
 105 110 115

gac tgc aag gac cag gcc gac tgc tgc atc aac tac agt tct gtg tgt 559
 Asp Cys Lys Asp Gln Gly Asp Cys Cys Ile Asn Tyr Ser Ser Val Cys
 120 125 130

caa ggt gag aaa agt tgg gta gaa gaa cca tgt gag agc att aat gag 607
 Gln Gly Glu Lys Ser Trp Val Glu Glu Pro Cys Glu Ser Ile Asn Glu
 135 140 145

cca cag tgc cca gca ggg ttt gaa acg cct cct acc ctc tta ttt tct 655
 Pro Gln Cys Pro Ala Gly Phe Glu Thr Pro Pro Thr Leu Leu Phe Ser
 150 155 160

ttg gat gga ttc agg gca gaa tat tta cac act tgg ggt gga ctt ctt 703
 Leu Asp Gly Phe Arg Ala Glu Tyr Leu His Thr Trp Gly Gly Leu Leu
 165 170 175 180

cct gtt att agc aaa cta aaa aaa tgt gga aca tat act aaa aac atg 751
 Pro Val Ile Ser Lys Leu Lys Lys Cys Gly Thr Tyr Thr Lys Asn Met
 185 190 195

aga ccg gta tat cca aca aaa act ttc ccc aat cac tac agc att gtc 799
 Arg Pro Val Tyr Pro Thr Lys Thr Phe Pro Asn His Tyr Ser Ile Val
 200 205 210

acc gga ttg tat cca gaa tct cat gcc ata atc gac aat aaa atg tat 847
 Thr Gly Leu Tyr Pro Glu Ser His Gly Ile Ile Asp Asn Lys Met Tyr
 215 220 225

gat ccc aaa atg aat gct tcc ttt tca ctt aaa agt aaa gag aaa ttt 895
 Asp Pro Lys Met Asn Ala Ser Phe Ser Leu Lys Ser Lys Glu Lys Phe

-continued

230	235	240	
aat cct gag tgg tac aaa gga gaa cca att tgg gtc aca gct aag tat			943
Asn Pro Glu Trp Tyr Lys Gly Glu Pro Ile Trp Val Thr Ala Lys Tyr			
245	250	255	260
caa ggc ctc aag tct ggc aca ttt ttc tgg cca gga tca gat gtg gaa			991
Gln Gly Leu Lys Ser Gly Thr Phe Phe Trp Pro Gly Ser Asp Val Glu			
265	270		275
att aac gga att ttc cca gac atc tat aaa atg tat aat ggt tca gta			1039
Ile Asn Gly Ile Phe Pro Asp Ile Tyr Lys Met Tyr Asn Gly Ser Val			
280	285		290
cca ttt gaa gaa agg att tta gct gtt ctt cag tgg cta cag ctt cct			1087
Pro Phe Glu Glu Arg Ile Leu Ala Val Leu Gln Trp Leu Gln Leu Pro			
295	300		305
aaa gat gaa aga cca cac ttt tac act ctg tat tta gaa gaa cca gat			1135
Lys Asp Glu Arg Pro His Phe Tyr Thr Leu Tyr Leu Glu Glu Pro Asp			
310	315		320
tct tca ggt cat tca tat gga cca gtc agc agt gaa gtc atc aaa gcc			1183
Ser Ser Gly His Ser Tyr Gly Pro Val Ser Ser Glu Val Ile Lys Ala			
325	330		340
ttg cag agg gtt gat ggt atg gtt ggt atg ctg atg gat ggt ctg aaa			1231
Leu Gln Arg Val Asp Gly Met Val Gly Met Leu Met Asp Gly Leu Lys			
345	350		355
gag ctg aac ttg cac aga tgc ctg aac ctc atc ctt att tca gat cat			1279
Glu Leu Asn Leu His Arg Cys Leu Asn Leu Ile Leu Ile Ser Asp His			
360	365		370
ggc atg gaa caa ggc agt tgt aag aaa tac ata tat ctg aat aaa tat			1327
Gly Met Glu Gln Gly Ser Cys Lys Lys Tyr Ile Tyr Leu Asn Lys Tyr			
375	380		385
ttg ggg gat gtt aaa aat att aaa gtt atc tat gga cct gca gct cga			1375
Leu Gly Asp Val Lys Asn Ile Lys Val Ile Tyr Gly Pro Ala Ala Arg			
390	395		400
ttg aga ccc tct gat gtc cca gat aaa tac tat tca ttt aac tat gaa			1423
Leu Arg Pro Ser Asp Val Pro Asp Lys Tyr Tyr Ser Phe Asn Tyr Glu			
405	410		420
ggc att gcc cga aat ctt tct tgc cgg gaa cca aac cag cac ttc aaa			1471
Gly Ile Ala Arg Asn Leu Ser Cys Arg Glu Pro Asn Gln His Phe Lys			
425	430		435
cct tac ctg aaa cat ttc tta cct aag cgt ttg cac ttt gct aag agt			1519
Pro Tyr Leu Lys His Phe Leu Pro Lys Arg Leu His Phe Ala Lys Ser			
440	445		450
gat aga att gag ccc ttg aca ttc tat ttg gac cct cag tgg caa ctt			1567
Asp Arg Ile Glu Pro Leu Thr Phe Tyr Leu Asp Pro Gln Trp Gln Leu			
455	460		465
gca ttg aat ccc tca gaa agg aaa tat tgt gga agt gga ttt cat ggc			1615
Ala Leu Asn Pro Ser Glu Arg Lys Tyr Cys Gly Ser Gly Phe His Gly			
470	475		480
tct gac aat gta ttt tca aat atg caa gcc ctc ttt gtt ggc tat gga			1663
Ser Asp Asn Val Phe Ser Asn Met Gln Ala Leu Phe Val Gly Tyr Gly			
485	490		500
cct gga ttc aag cat ggc att gag gct gac acc ttt gaa aac att gaa			1711
Pro Gly Phe Lys His Gly Ile Glu Ala Asp Thr Phe Glu Asn Ile Glu			
505	510		515
gtc tat aac tta atg tgt gat tta ctg aat ttg aca ccg gct cct aat			1759
Val Tyr Asn Leu Met Cys Asp Leu Leu Asn Leu Thr Pro Ala Pro Asn			
520	525		530
aac gga act cat gga agt ctt aac cac ctt cta aag aat cct gtt tat			1807
Asn Gly Thr His Gly Ser Leu Asn His Leu Leu Lys Asn Pro Val Tyr			
535	540		545
acg cca aag cat ccc aaa gaa gtg cac ccc ctg gta cag tgc ccc ttc			1855

Thr 550	Pro 550	Lys	His	Pro	Lys	Glu 555	Val	His	Pro	Leu	Val 560	Gln	Cys	Pro	Phe	
aca Thr 565	aga Arg	aac Asn	ccc Pro	aga Arg	gat Asp 570	aac Asn	ctt Leu	ggc Gly	tgc Cys	tca Ser 575	tgt Cys	aac Asn	cct Pro	tcg Ser	att Ile 580	1903
ttg Leu	ccg Pro	att Ile	gag Glu	gat Asp 585	ttt Phe	caa Gln	aca Thr	cag Gln	ttc Phe 590	aat Asn	ctg Leu	act Thr	gtg Val	gca Ala 595	gaa Glu	1951
gag Glu	aag Lys	att Ile	att Ile 600	aag Lys	cat His	gaa Glu	act Thr	tta Leu 605	ccc Pro	tat Tyr	gga Gly	aga Arg	cct Pro 610	aga Arg	gtt Val	1999
ctc Leu	cag Gln	aag Lys	gaa Glu 615	aac Asn	acc Thr	atc Ile	tgt Cys 620	ctt Leu	ctt Leu	tcc Ser	cag Gln	cac His	cag Gln 625	ttt Phe	atg Met	2047
agt Ser	gga Gly	tac Tyr	agc Ser	caa Gln	gac Asp 635	atc Ile	tta Leu	atg Met	ccc Pro	ctt Leu	tgg Trp 640	aca Thr	tcc Ser	tat Tyr	acc Thr	2095
gtg Val 645	gac Asp	aga Arg	aat Asn	gac Asp	agt Ser 650	ttc Phe	tct Ser	acg Thr	gaa Glu	gac Phe 655	ttc Phe	tcc Ser	aac Asn	tgt Cys	ctg Leu 660	2143
tac Tyr	cag Gln	gac Asp	ttt Phe	aga Arg 665	att Ile	cct Pro	ctt Leu	agt Ser	cct Pro 670	gtc Val	cat His	aaa Lys	tgt Cys	tca Ser 675	ttt Phe	2191
tat Tyr	aaa Lys	aat Asn	aac Asn 680	acc Thr	aaa Lys	gtg Val	agt Ser	tac Tyr 685	ggg Gly	ttc Phe	ctc Leu	tcc Ser	cca Pro 690	caa Pro	caa Gln	2239
cta Leu	aat Asn	aaa Lys 695	aat Asn	tca Ser	agt Ser	gga Gly	ata Ile 700	tat Tyr	tct Ser	gaa Glu	gct Ala	ttg Leu 705	ctt Leu	act Thr	aca Thr	2287
aat Asn	ata Ile 710	gtg Val	cca Pro	atg Met	tac Tyr	cag Gln 715	agt Ser	ttt Phe	caa Gln	gtt Val 720	ata Ile	tgg Trp	cgc Arg	tac Tyr	ttt Phe	2335
cat His 725	gac Asp	acc Thr	cta Leu	ctg Leu	cga Arg 730	aag Lys	tat Tyr	gct Ala	gaa Glu	gaa Glu 735	aga Arg	aat Asn	ggg Gly	gtc Val	aat Asn 740	2383
gtc Val	gtc Val	agt Ser	ggg Gly	cct Pro 745	gtg Val	ttt Phe	gac Asp	ttt Phe	gat Asp 750	tat Tyr	gat Asp	gga Gly	cgt Arg	tgt Cys 755	gat Asp	2431
tcc Ser	tta Leu	gag Glu	aat Asn 760	ctg Leu	agg Arg	caa Gln	aaa Lys	aga Arg 765	aga Arg	gtc Val	atc Ile	cgt Arg	aac Asn 770	caa Gln	gaa Glu	2479
att Ile	ttg Leu	att Ile 775	cca Pro	act Thr	cac His	ttc Phe	ttt Phe 780	att Ile	gtg Val	cta Leu	aca Thr	agc Ser	tgt Cys 785	aaa Lys	gat Asp	2527
aca Thr	tct Ser	cag Gln	acg Thr	cct Pro	ttg Leu	cac His 795	tgt Cys	gaa Glu	aac Asn	cta Leu	gac Asp 800	acc Thr	tta Leu	gct Ala	ttc Phe	2575
att Ile 805	ttg Leu	cct Pro	cac His	agg Arg	act Thr 810	gat Asp	aac Asn	agc Ser	gag Glu	agc Ser 815	tgt Cys	gtg Val	cat His	ggg Gly	aag Lys 820	2623
cat His	gac Asp	tcc Ser	tca Ser	tgg Trp 825	gtt Val	gaa Glu	gaa Glu	ttg Leu	tta Leu 830	atg Met	tta Leu	cac His	aga Arg	gca Ala 835	cgg Arg	2671
atc Ile	aca Thr	gat Asp	gtt Val 840	gag Glu	cac His	atc Ile	act Thr	gga Gly 845	ctc Leu	agc Ser	ttc Phe	tat Tyr	caa Gln 850	caa Gln	aga Arg	2719
aaa Lys	gag Glu	cca Pro 855	gtt Val	tca Ser	gac Asp	att Ile	tta Leu 860	aag Lys	ttg Leu	aaa Lys	aca Thr	cat His 865	ttg Leu	cca Pro	acc Thr	2767

-continued

ttt agc caa gaa gac tga tatgtttttt atccccaac accatgaatc 2815
 Phe Ser Gln Glu Asp
 870

tttttgagag aaccttatat tttatatagt cctctagcta cactattgca ttgttcagaa 2875
 actgtcgacc agagttagaa cggagccctc ggtgatgcgg acatctcagg gaaacttgcg 2935
 tactcagcac agcagtgagg agtgttcctg ttgaatcttg cacatatattg aatgtgtaag 2995
 cattgtatac attgatcaag ttcgggggaa taaagacaga ccacacctaa aactgccttt 3055
 ctgcttctct taaaggagaa gtactgtga acattgtctg gataccagat attgaaatct 3115
 ttcttactat tggtataaaa ccttgatggc attgggcata cagtagactt atagtagggt 3175
 tggggtagcc catgttatgt gactatcttt atgagaattt taaagtgtt ctggatatct 3235
 ttttaacttg agtttcattt cttttcattg taatcaaaaa aaaaattaac agaagccaaa 3295
 atactctga gacctgttt caatcttgc tgtatatccc ctcaaaatcc aagttattaa 3355
 tcttatgtgt tttcttttta attttttgat tggatttctt tagatttaat ggttcaaatg 3415
 agttcaactt tgaggagcga tctttgaata tacttaccta ttataaaatc ttactttgta 3475
 ttgtattta a 3486

<210> SEQ ID NO 4
 <211> LENGTH: 873
 <212> TYPE: PRT
 <213> ORGANISM: H. sapiens

<400> SEQUENCE: 4

Met Asp Val Gly Glu Glu Pro Leu Glu Lys Ala Ala Arg Ala Arg Thr
 1 5 10 15
 Ala Lys Asp Pro Asn Thr Tyr Lys Val Leu Ser Leu Val Leu Ser Val
 20 25 30
 Cys Val Leu Thr Thr Ile Leu Gly Cys Ile Phe Gly Leu Lys Pro Ser
 35 40 45
 Cys Ala Lys Glu Val Lys Ser Cys Lys Gly Arg Cys Phe Glu Arg Thr
 50 55 60
 Phe Gly Asn Cys Arg Cys Asp Ala Ala Cys Val Glu Leu Gly Asn Cys
 65 70 75 80
 Cys Leu Asp Tyr Gln Glu Thr Cys Ile Glu Pro Glu His Ile Trp Thr
 85 90 95
 Cys Asn Lys Phe Arg Cys Gly Glu Lys Arg Leu Thr Arg Ser Leu Cys
 100 105 110
 Ala Cys Ser Asp Asp Cys Lys Asp Gln Gly Asp Cys Cys Ile Asn Tyr
 115 120 125
 Ser Ser Val Cys Gln Gly Glu Lys Ser Trp Val Glu Glu Pro Cys Glu
 130 135 140
 Ser Ile Asn Glu Pro Gln Cys Pro Ala Gly Phe Glu Thr Pro Pro Thr
 145 150 155 160
 Leu Leu Phe Ser Leu Asp Gly Phe Arg Ala Glu Tyr Leu His Thr Trp
 165 170 175
 Gly Gly Leu Leu Pro Val Ile Ser Lys Leu Lys Lys Cys Gly Thr Tyr
 180 185 190
 Thr Lys Asn Met Arg Pro Val Tyr Pro Thr Lys Thr Phe Pro Asn His
 195 200 205
 Tyr Ser Ile Val Thr Gly Leu Tyr Pro Glu Ser His Gly Ile Ile Asp
 210 215 220
 Asn Lys Met Tyr Asp Pro Lys Met Asn Ala Ser Phe Ser Leu Lys Ser

-continued

225	230	235	240
Lys Glu Lys Phe Asn Pro Glu Trp Tyr Lys Gly Glu Pro Ile Trp Val			
	245	250	255
Thr Ala Lys Tyr Gln Gly Leu Lys Ser Gly Thr Phe Phe Trp Pro Gly			
	260	265	270
Ser Asp Val Glu Ile Asn Gly Ile Phe Pro Asp Ile Tyr Lys Met Tyr			
	275	280	285
Asn Gly Ser Val Pro Phe Glu Glu Arg Ile Leu Ala Val Leu Gln Trp			
	290	295	300
Leu Gln Leu Pro Lys Asp Glu Arg Pro His Phe Tyr Thr Leu Tyr Leu			
	305	310	315
Glu Glu Pro Asp Ser Ser Gly His Ser Tyr Gly Pro Val Ser Ser Glu			
	325	330	335
Val Ile Lys Ala Leu Gln Arg Val Asp Gly Met Val Gly Met Leu Met			
	340	345	350
Asp Gly Leu Lys Glu Leu Asn Leu His Arg Cys Leu Asn Leu Ile Leu			
	355	360	365
Ile Ser Asp His Gly Met Glu Gln Gly Ser Cys Lys Lys Tyr Ile Tyr			
	370	375	380
Leu Asn Lys Tyr Leu Gly Asp Val Lys Asn Ile Lys Val Ile Tyr Gly			
	385	390	395
Pro Ala Ala Arg Leu Arg Pro Ser Asp Val Pro Asp Lys Tyr Tyr Ser			
	405	410	415
Phe Asn Tyr Glu Gly Ile Ala Arg Asn Leu Ser Cys Arg Glu Pro Asn			
	420	425	430
Gln His Phe Lys Pro Tyr Leu Lys His Phe Leu Pro Lys Arg Leu His			
	435	440	445
Phe Ala Lys Ser Asp Arg Ile Glu Pro Leu Thr Phe Tyr Leu Asp Pro			
	450	455	460
Gln Trp Gln Leu Ala Leu Asn Pro Ser Glu Arg Lys Tyr Cys Gly Ser			
	465	470	475
Gly Phe His Gly Ser Asp Asn Val Phe Ser Asn Met Gln Ala Leu Phe			
	485	490	495
Val Gly Tyr Gly Pro Gly Phe Lys His Gly Ile Glu Ala Asp Thr Phe			
	500	505	510
Glu Asn Ile Glu Val Tyr Asn Leu Met Cys Asp Leu Leu Asn Leu Thr			
	515	520	525
Pro Ala Pro Asn Asn Gly Thr His Gly Ser Leu Asn His Leu Leu Lys			
	530	535	540
Asn Pro Val Tyr Thr Pro Lys His Pro Lys Glu Val His Pro Leu Val			
	545	550	555
Gln Cys Pro Phe Thr Arg Asn Pro Arg Asp Asn Leu Gly Cys Ser Cys			
	565	570	575
Asn Pro Ser Ile Leu Pro Ile Glu Asp Phe Gln Thr Gln Phe Asn Leu			
	580	585	590
Thr Val Ala Glu Glu Lys Ile Ile Lys His Glu Thr Leu Pro Tyr Gly			
	595	600	605
Arg Pro Arg Val Leu Gln Lys Glu Asn Thr Ile Cys Leu Leu Ser Gln			
	610	615	620
His Gln Phe Met Ser Gly Tyr Ser Gln Asp Ile Leu Met Pro Leu Trp			
	625	630	635
Thr Ser Tyr Thr Val Asp Arg Asn Asp Ser Phe Ser Thr Glu Asp Phe			
	645	650	655

-continued

Ser Asn Cys Leu Tyr Gln Asp Phe Arg Ile Pro Leu Ser Pro Val His
 660 665 670
 Lys Cys Ser Phe Tyr Lys Asn Asn Thr Lys Val Ser Tyr Gly Phe Leu
 675 680 685
 Ser Pro Pro Gln Leu Asn Lys Asn Ser Ser Gly Ile Tyr Ser Glu Ala
 690 695 700
 Leu Leu Thr Thr Asn Ile Val Pro Met Tyr Gln Ser Phe Gln Val Ile
 705 710 715 720
 Trp Arg Tyr Phe His Asp Thr Leu Leu Arg Lys Tyr Ala Glu Glu Arg
 725 730 735
 Asn Gly Val Asn Val Val Ser Gly Pro Val Phe Asp Phe Asp Tyr Asp
 740 745 750
 Gly Arg Cys Asp Ser Leu Glu Asn Leu Arg Gln Lys Arg Arg Val Ile
 755 760 765
 Arg Asn Gln Glu Ile Leu Ile Pro Thr His Phe Phe Ile Val Leu Thr
 770 775 780
 Ser Cys Lys Asp Thr Ser Gln Thr Pro Leu His Cys Glu Asn Leu Asp
 785 790 795 800
 Thr Leu Ala Phe Ile Leu Pro His Arg Thr Asp Asn Ser Glu Ser Cys
 805 810 815
 Val His Gly Lys His Asp Ser Ser Trp Val Glu Glu Leu Leu Met Leu
 820 825 830
 His Arg Ala Arg Ile Thr Asp Val Glu His Ile Thr Gly Leu Ser Phe
 835 840 845
 Tyr Gln Gln Arg Lys Glu Pro Val Ser Asp Ile Leu Lys Leu Lys Thr
 850 855 860
 His Leu Pro Thr Phe Ser Gln Glu Asp
 865 870

<210> SEQ ID NO 5
 <211> LENGTH: 646
 <212> TYPE: DNA
 <213> ORGANISM: H. sapiens
 <220> FEATURE:
 <221> NAME/KEY: promoter
 <222> LOCATION: (0)...(0)
 <223> OTHER INFORMATION: PC-1 promoter sequence
 <220> FEATURE:
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (1)...(646)
 <223> OTHER INFORMATION: n = A,T,C or G

<400> SEQUENCE: 5

aaaccaacgt agnagcgtgg gaatcgaaat atccttaggt gtgttcagta tatgtgaacc 60
 cacgtatttt aagtggacga ttctctcttc agagtaccgt aggtagtggg ggacggggcg 120
 cagaggggga gaaacagaaa gtcgccttcc tccatgggtc atttgcatth ccatccagaa 180
 actcacaggt cgaccccaag actccactct ctccccctt tgagaagccg gaccggcatc 240
 ggcggtgca tccttctcct cctccccgct ctattttggg gcccatgat ctcatgcctt 300
 ctgcagacca cacgtgcaa ttccagccca gccgcgccc cgaggccacg cagggcgatt 360
 cctgcaagtg tcgggagggg ggccggggcg cggggagggg acggcctggg gggaagtta 420
 agacacgccc acgtaaggga cccaaaataa ccgacacaca gagtgccga aatcagacag 480
 gaagccaaat aatccggggc gttgagtcgc ttgcccctga ctgagagagc cggtgttagg 540
 gcggggagcc aaggatctga ccgcgagggg cgggcgccgc ggggaggggc ggggcggggc 600

-continued

gggcggcgcg gggcctatta aaggcgcgcg ggggcagcgg ggcggg 646

<210> SEQ ID NO 6
<211> LENGTH: 350
<212> TYPE: DNA
<213> ORGANISM: H. sapiens
<220> FEATURE:
<221> NAME/KEY: 3'UTR
<222> LOCATION: (0)...(0)
<223> OTHER INFORMATION: allele "A"

<400> SEQUENCE: 6

agccaagaag actgatatgt tttttatccc caaacacccat gaatcttttt gagagaacct 60
tatattttat atagtcctct agctacacta ttgcattggt cagaaactgt cgaccagagt 120
tagaacggag ccctcgggtga tgcggacatc tcagggaaac ttgcgtactc agcacagcag 180
tggagagtgt tcctgttgaa tcttgacatc atttgaatgt gtaagcattg tatacattga 240
tcaagttcgg gggaataaag acagaccaca cctaaaactg cctttctgct tctcttaaag 300
gagaagtagc tgtgaacatt gtctggatag cagatatttg aatctttctt 350

<210> SEQ ID NO 7
<211> LENGTH: 350
<212> TYPE: DNA
<213> ORGANISM: H. sapiens
<220> FEATURE:
<221> NAME/KEY: 3'UTR
<222> LOCATION: (0)...(0)
<223> OTHER INFORMATION: allele "P"

<400> SEQUENCE: 7

agccaagaag actgatatgt tttttatccc caaacacccat gaatcttttt gagagaacct 60
tatattttat atagtcctct agctacacta ttgcattggt cagaaactgt cgaccagagt 120
tagaacagag ccctccgtga tgcggacatc tcagggaaac ttgcgtactc agcacagtag 180
tggagagtgt tcctgttgaa tcttgacatc atttgaatgt gtaagcattg tatacattga 240
tcaagttcgg gggaataaag acagaccaca cctaaaactg cctttctgct tctcttaaag 300
gagaagtagc tgtgaacatt gtctggatag cagatatttg aatctttctt 350

<210> SEQ ID NO 8
<211> LENGTH: 350
<212> TYPE: DNA
<213> ORGANISM: H. sapiens
<220> FEATURE:
<221> NAME/KEY: 3'UTR
<222> LOCATION: (0)...(0)
<223> OTHER INFORMATION: allele "N"

<400> SEQUENCE: 8

agccaagaag actgatatgt tttttatccc caaacacccat gaatcttttt gagagaacct 60
tatattttat atagtcctct agctacacta ttgcattggt cagaaactgt cgaccagagt 120
tagaacagag ccctcgggtga tgcggacatc tcagggaaac ttgcgtactc agcacagtag 180
tggagagtgt tcctgttgaa tcttgacatc atttgaatgt gtaagcattg tatacattga 240
tcaagttcgg gggaataaag acagaccaca cctaaaactg cctttctgct tctcttaaag 300
gagaagtagc tgtgaacatt gtctggatag cagatatttg aatctttctt 350

<210> SEQ ID NO 9
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: H. sapiens

-continued

<400> SEQUENCE: 9

ctgtgttcac ttggacatg ttg 23

<210> SEQ ID NO 10

<211> LENGTH: 22

<212> TYPE: DNA

<213> ORGANISM: H. sapiens

<400> SEQUENCE: 10

gacgttggaa gataccaggt tg 22

<210> SEQ ID NO 11

<211> LENGTH: 109

<212> TYPE: DNA

<213> ORGANISM: H. sapiens

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (1)...(109)

<223> OTHER INFORMATION: n = A,T,C or G

<400> SEQUENCE: 11

ctctcgctgg taggtccgcg gccaggcccc ggcgcccggg agggctggga atacngggag 60

ggcgcgcgcg agctcctgcg ctctcagcgc actcagcacc gggcacgga 109

<210> SEQ ID NO 12

<211> LENGTH: 279

<212> TYPE: DNA

<213> ORGANISM: H. sapiens

<400> SEQUENCE: 12

tgagctccac cgggccggcg gccgctctag aactagtggg tcatgccact gtaccctagc 60

ctgggtaaca gagtaagaca ctatctctaa aaataaaaaa taagataaaa tattttttaa 120

aaaagaaacc atgtaatttt ctcttttctc cctacaggta ttgagaaggt aattaggtgt 180

gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtatgtgtgt gcacagcctt attaagaatg 240

tgattgaggt aaacattatc tcctattccc aaggggtac 279

<210> SEQ ID NO 13

<211> LENGTH: 243

<212> TYPE: DNA

<213> ORGANISM: H. sapiens

<400> SEQUENCE: 13

agatttttgc cttactttat taccocatct gtattttcta aagtagtatt tgaacctagt 60

gtacacctaa cttagtgtga ttctgtgatg tttactttga attatataat gattagaaac 120

atctgactta tcgttcaatt ttttcagtta accaggtaag gatgagcagg gaaaaaagtg 180

gagttatggt cattaggaaa agatccacta gttctagagc ggccgccacg cccgggtggag 240

ctt 243

<210> SEQ ID NO 14

<211> LENGTH: 231

<212> TYPE: DNA

<213> ORGANISM: H. sapiens

<220> FEATURE:

<221> NAME/KEY: misc_feature

<222> LOCATION: (1)...(231)

<223> OTHER INFORMATION: n = A,T,C or G

<400> SEQUENCE: 14

-continued

cgcggcggcc gttctagaac tagtgatca tactcacgaa gacagcaatt ctgtgttcac	60
tttgacatg ttgaattga gacataaac acattttgct gatgtttgt tctagaacat	120
agtcaaggtc aggtgctcgt tgggctctgc agcaacctgg tatcttccaa cctcttaacg	180
gggcntaca taagtgttat cttttatatt aagantcatg gctattgggc c	231

<210> SEQ ID NO 15
 <211> LENGTH: 313
 <212> TYPE: DNA
 <213> ORGANISM: H. sapiens

<400> SEQUENCE: 15

aatctgttca catactttgt ttgtggaatc tgtcttaatg tgtctcacia gcatcacaat	60
tattattact gttaagtgtg ttcattttat tttcttgaaa atatttttagg tgagaagcag	120
ggtaagatta tattctgagg tattaatttt tttcttttta gaagtacagc atcatttttt	180
tctttccaaa ttaagatgat aaaaataata aaatcactgg tttattaaac attacagggt	240
gagtatcctt tatccaaaat gtttggtatg agaactgttt tggattttgg acttttttgg	300
attttgcaat att	313

<210> SEQ ID NO 16
 <211> LENGTH: 313
 <212> TYPE: DNA
 <213> ORGANISM: H. sapiens

<400> SEQUENCE: 16

ccgcagcccg ggggatcaca cagaccttag tggaaaatct tcaactggacc tgtccaaga	60
agggggtaca tcttcattgg atatgtcttg tctttgcttc tttaaacatt ttttttctt	120
tttcattacc caggtttgaa actaagttag taacttcaga gtttactgct ggaatatcac	180
catttcagt agattgacta ggcaggcagt ctttcttgga aaagtactgg cagaacctaa	240
ctgtttcact aaacttttct aatgggcaaa gtagttgaac cttgtgtagg gcgccttacc	300
tttaataatg tga	313

<210> SEQ ID NO 17
 <211> LENGTH: 382
 <212> TYPE: DNA
 <213> ORGANISM: H. sapiens

<400> SEQUENCE: 17

taagagaaaa atgaagtcac cttaagatt ggatttgtat ccacagtgtt gctttataat	60
tcacctgaa tttttatctg attaaaatcc ctctgggta atttttttta cgtgatttag	120
actgctgtgg taccactgct aaatgaggta agccaattgt cagatgtatt taantaacat	180
gtttattttt ttcccttcta gaaaaatgtt caccgtaagc tctgcatttc aacttctatc	240
tgttgaaga agtgagatgg gattgtaaca ttttttgagg gaatagattt aagataaaag	300
aaaaacaact tattttccaa taggtagtta agtaaggaaa cccaggttct gatctttgct	360
ctgccacaaa ctagctgtgg ct	382

<210> SEQ ID NO 18
 <211> LENGTH: 312
 <212> TYPE: DNA
 <213> ORGANISM: H. sapiens

<400> SEQUENCE: 18

actacataaa atcttaagag gttgcgtttt gccattacct gatTTTTTTTg tttttctttc	60
--	----

-continued

cttaaaactta ttataattcc atgtagcttc agttatcggt ttctttttga tgattttttt 120
ctgtgaatgt atttaacatt aagtaaacac aacttgcata taatctgttt tatctttttt 180
agggattaac cagtgagttc ttgttttttc tactaaaata gttaattatt ctcactctatt 240
tcaatcagag taaaataacc agattctcta gagcttttaa taactgattt catttagtgt 300
gtctgtggcc at 312

<210> SEQ ID NO 19
<211> LENGTH: 382
<212> TYPE: DNA
<213> ORGANISM: H. sapiens

<400> SEQUENCE: 19

taatctctga ctatttaata tgttggtgct gcttaagagt catattacat gattattgtc 60
gtctaagtgc tgaagcttgt tgacctaaa agcattctag cactagagag gaatgcattg 120
gtgtgggtatg aaacatact ttcctaagag atgaatgttg catgatttct taattttcct 180
tcattttctg ctccagattt ggaatgggta tgtgaaatga attttttcta ggatctgtaa 240
tatagaacag cttattctta tgtaactctc tttttattga atcctgagct ttagcatttg 300
agtgatatgt tggctgaaaa atgagaactg aagaactctt tctcaagag ttagataga 360
tggtaatgg acagtaaac ta 382

<210> SEQ ID NO 20
<211> LENGTH: 307
<212> TYPE: DNA
<213> ORGANISM: H. sapiens

<400> SEQUENCE: 20

gggaaaataa agttttcaaa taaaaccctt gatttcaaac acaatagatg cgaatatagca 60
tttactagct cttaatgaca ttttcaatga aaaaaactat attttacacc caaacaattg 120
tcagccatct tttatttttg ttgtttcttc atttttagttc agtaagatga aaggctctgta 180
ggcaattaat ttctattgta aatacttcgt ttgttagaaa tgatatacta ttttccocta 240
gactacaaca aaactttgct atttgctatg atgttttata tcgaaataaa ttcttttagta 300
aatgatac 307

<210> SEQ ID NO 21
<211> LENGTH: 329
<212> TYPE: DNA
<213> ORGANISM: H. sapiens

<400> SEQUENCE: 21

gaatttcaaa gctgtaaatt atttctcagt agaactgtta caccagtgtt ataaaattta 60
atccctatca attgaggaat ttttttttcc attctgtttt tcaatgtgtt cgtaaaatat 120
tacattttga tactgtttga tttagaccac acagtgagta agtacatttt tctcagtaat 180
tatttcatta aaccagtcga tcaaactgaa cctcgctttg aaggaggctg ctgagaccatt 240
ttataagatt ctatcatttc tggaaaaagc aagtattata cacaatatta ctaaatataa 300
ggatgcactt taaacaaaat aagagttgg 329

<210> SEQ ID NO 22
<211> LENGTH: 381
<212> TYPE: DNA
<213> ORGANISM: H. sapiens

-continued

<400> SEQUENCE: 22

```
gtcttagttt aatgtgaatc agctcattgt agttgcatcc actggcccaa atctatcaat    60
ctgtcgggtct ttcttttctt ctttgtttct ttcttttttt ttttttttaa cagagatagc    120
tttatgtata aatagccatt agtgtggaag gtatcacatg aggttggtct tccattctt    180
aggtcatcat catggtaatc tgaatttgca ttatttactc ttcaggataa agggctgaag    240
aaagtttact tgatgggttc ccaatttttt ctgaatgttg tagttaattc ttttttaaaa    300
atgtagtttc ttatggacag tctttaggaa aaaaatacat taaatataaa atataagtga    360
aacacagaat tcacagaaac c                                         381
```

<210> SEQ ID NO 23

<211> LENGTH: 382

<212> TYPE: DNA

<213> ORGANISM: H. sapiens

<400> SEQUENCE: 23

```
gattttgaaa aaagtgaagt gataggtaca gctgaaattc tgtcttacct atcagatctt    60
caactaatat gagtgctaca cccatgttta acgaatttaa ccttggaagt gaaagaagtt    120
ctgctctgca tattaaattt ttgtttaaag ttacagcatg ttttgggatt tttttttct    180
cctaggcatg gtactattca tgtaagtata tctctgtgat aactttgaat atggtcatat    240
taagaatacc ttccttttag ccgggcacag tggctcatgc ctgtaatcgc agcactttgg    300
gaggccaaag tgggtggtca cctgaggtca ggagttcgag accagcctgg ccaacatggt    360
gaaaccctaa aaatacatat ac                                         382
```

<210> SEQ ID NO 24

<211> LENGTH: 361

<212> TYPE: DNA

<213> ORGANISM: H. sapiens

<400> SEQUENCE: 24

```
gatccaaact ctgcatttaa ataccaaggc aggtttttaa gagttcattt aagtcattac    60
attgtagcca ctgaaaggaa ttagacagac ctttagggat ctgacattct atatttttgt    120
attatgtttt aattatagta tacaatcaac tattaattct tatgtttgtt cccctccagt    180
taactatgaa ggcattggcc gaaatctttc tgtgagtatc tttattttcc attatctagt    240
tatttttact tttgtataat atatatggag agaaaagttt cagcatctat tattgggatt    300
gaaggattag aatattttag taatctgggc caacetggaa atgctgtgta gtttaaagat    360
c                                         361
```

<210> SEQ ID NO 25

<211> LENGTH: 384

<212> TYPE: DNA

<213> ORGANISM: H. sapiens

<400> SEQUENCE: 25

```
ctgatgaaat gtttgtgaaa aaaaatttca tatgaagtta gaaagcaatt tcaagaaaag    60
ttgacacttt ttatagatat tagggaaata tctttcccta ataaatatct ttccctaaaa    120
aagttgacac ttttttagat attagggaaa taatagtttt tctttgctgt ttgcaatttc    180
agtgccgggg cattgtaagt tctgacagtc tcccaggtaa acttagtctg atcggttagt    240
gattcagggt aaccattggg ccctttctaa caatattgtt atgtgaaaac tgtataagta    300
tgattctctt cactctaacc caggatttct aatgtcggcc tatggatgtt tgagttagat    360
```

-continued

aattctttgt tgtggagagc tgtc 384

<210> SEQ ID NO 26
<211> LENGTH: 381
<212> TYPE: DNA
<213> ORGANISM: H. sapiens

<400> SEQUENCE: 26

aaaagataga ggtgacttct taatgctttt caaagccagg tggttttatt taccgttggtg 60
ttggtttaac aaaatagtta catacttttt aatcaatgaa aataatgta tgattatcaa 120
ttatgtttta tgaaggact ttacattttt aattcatata tgtcaacatt aggaatgcaa 180
gtgagtaaac ctattatact taattggatt aaatctaaag aaaaaatgat atgcaaagtt 240
ttagacttga aaacatactg tgattatatg tcttgaatga gaattaatgg aacatacttt 300
cataaagcta tttttctttg aacattaaag aattttgtta aagttttata ttcattggct 360
attactaaaa agtcaaaaaa c 381

<210> SEQ ID NO 27
<211> LENGTH: 383
<212> TYPE: DNA
<213> ORGANISM: H. sapiens

<400> SEQUENCE: 27

aaaactaaga gacctatcct agatgtcctt agattatgtg tgtgataggg ttaaaactat 60
atttcccaca aagtcactg agcgtggtag ttttcctctt atcttatcat aaccagtttg 120
tatatgtaca atgtggataa cagaattttt gggaccaact tgtagacagc tgaatgcac 180
tgataaactt ccttttctgg ccatctaggc cctgtgtggt aagtgtgaac aggtgccttt 240
tttccttctt gaaaatagac ctgaaatagg attatcaaaa gcaggtcaca ttgtaggcaa 300
ctttgtggag atgatggga ggcaagacag atttttacct tcttcctgac tctcagactc 360
actgaagaaa tgtggggaac atg 383

<210> SEQ ID NO 28
<211> LENGTH: 384
<212> TYPE: DNA
<213> ORGANISM: H. sapiens

<400> SEQUENCE: 28

catatcagta tttctattaa aaataaccta gtcttaaata ctctaaaacc caagagagtt 60
ttatactttt attttagtta aagagtaaat gactcatgta tttggtttta aaaaagttaa 120
gatcatggca caagtctact atttgtttga ttgaaacat ctaagtaact ctaccatctt 180
gaaattatgc agatttactt cggtaagtat cgtcaagaag tttggtccag tatgtatggt 240
ttgatagcac cctctgcata gcatgtgctg taaaaaact taataatcaa attagaattt 300
aggagtgggg gtaggtaaac atatgtttta attctagggg gcgcatgtaa atcttttgtg 360
atatatcttt tctctttcta gttt 384

<210> SEQ ID NO 29
<211> LENGTH: 339
<212> TYPE: DNA
<213> ORGANISM: H. sapiens

<400> SEQUENCE: 29

gtgaaagagc aacactcttg ccttgaaaga gaaaaaaaaa tccactaata caagactatc 60

-continued

ataaatgatac tttgttctat gttggaataa tcaatctata gcggtctatg ttacaaaatt	120
taaaacatgt ctctcagtc ttacaaatag ttttataacc ttttttcaga ttttgccgaa	180
ggtaaggcat gctacacact caagctcgga atgtgaagca ggcattttct catcagtgtg	240
aaatgcagag aactggccttg ggggtattat ttgagaataa ccaataaaaat aaagggaggt	300
ctggaggacc acctgatgaa acatagaggt ttctttgct	339

<210> SEQ ID NO 30
 <211> LENGTH: 327
 <212> TYPE: DNA
 <213> ORGANISM: H. sapiens

<400> SEQUENCE: 30

gtcttcttaa ttgtttatgc ttgtaccctt tgtaatcagt ttttttaata gttaaaagta	60
aatcttcaat ataattaagt agaggaaagg attagatgag tgtatcacac tatatattat	120
catataatgc acactaacta cttttatttt catcctgtga cccaagagaa gattagacag	180
aaatgcaagt atttgtcacc tctttatgtg tggccatttc aaattaatga ttaagcagaa	240
cattaaatgc atagtttctc actgttcacc ttggctttat actcagttcc cgcattagag	300
gaacactgaa gagggagtca gaaaaat	327

<210> SEQ ID NO 31
 <211> LENGTH: 427
 <212> TYPE: DNA
 <213> ORGANISM: H. sapiens

<400> SEQUENCE: 31

tttaatatgg taaagcattt ttacacttta gttagaaaaa aagatgaata tactagtagg	60
aaaatagggg aggacatgag ctgacagcta gagcttcata attttatgat gtggttcacc	120
tttaaatatt aataaagcaa ttttcttctc tgtgcctgat atctgagagt tcttctcatt	180
ttcgttcttc aggacaccac cacgtaagtt ttttctctc ctgaccttcc cttttctcct	240
ttttgttttc tttcttggtt ataaatccta ccatacatta tagggtaata tatatattac	300
ctattatata tatataaaat attacctatt ttatatatat attatatata taatatatat	360
aaagtatata tattactatt ttatatatat atagctatat atatatacct ttgtttattt	420
attgtga	427

<210> SEQ ID NO 32
 <211> LENGTH: 380
 <212> TYPE: DNA
 <213> ORGANISM: H. sapiens

<400> SEQUENCE: 32

ctcatcttga aaagacttct taaatatttt atttttgtaa aggacttgac caaacacata	60
acattttccc tcgaccctgt acttgggaaa gttttacagg ttttaagatgg tactcagcta	120
atttttaaaa atgctccccc aacctagaga aagtataatt tcctatgtta tttgtgaaga	180
atgaaaaagt tgcctctctt tctctttgta gaactattca aggtaaaata tgttaactct	240
atatttgata attttaatga atttgtgcac atataggcat aattcatatg tataggactt	300
atggtctaaa ttaaatgaat taataccaaa tacattctta aagggttaac tttgagaata	360
ctagtacaca aaaattctac	380

<210> SEQ ID NO 33
 <211> LENGTH: 384

-continued

```

<212> TYPE: DNA
<213> ORGANISM: H. sapiens

<400> SEQUENCE: 33
ctgggtgata tagcacgact ctgtctctaa acaaaaaaca aaacaaaacg aagactgaag    60
ccaaacttga ctttatcttt atttactata aatgctaatt ttgaatcatg gtgttaattt    120
atttcacacg tcaacatggt cccttgttct ttgaaacta cactggcttc tatcttggtt    180
cagttataga ggcagtaaga acatatttca ttactcttaa aaataggaat taccatccag    240
tagaaatggg attaccatcc agttgagtca acagaacctt tttatccag tgcgtatgt    300
ttatgtgtat gacactctcg actacacagg aagcctcttg aaatatctga ttaattttga    360
tgttttgctc aatgttcagt aaaa                                384

<210> SEQ ID NO 34
<211> LENGTH: 328
<212> TYPE: DNA
<213> ORGANISM: H. sapiens

<400> SEQUENCE: 34
gttcttatat ttaattattg gttggaattt gatttttata tgtattaaaa gcatgctcta    60
ctgaaatatt catcaaaagg aagatagtta tttctttctt aaaatgaata ttggcatgtt    120
ttacagaaaa atgtgttgga agtagctttt gtatatttac ttgcatgtt gaaaatctag    180
acatatgcat atttgtttat gtcacccatc tgacattaca gtgagagaaa gcacaactga    240
gtacacatgg acttcgaaat tataggatgc ttttaaatat gatcttttaa gatgacatat    300
ctttggggaa gactaccctg tctgcttt                                328

<210> SEQ ID NO 35
<211> LENGTH: 384
<212> TYPE: DNA
<213> ORGANISM: H. sapiens

<400> SEQUENCE: 35
aattaaacaa acatgcacgg tatgtattag aaggaaagct actcaagagg agagatgatg    60
cctaacaat catgtggcac gttccacttc agagctgaaa tctcgtaaat gattaaactg    120
gggagatgga gcacttatag aagtgaactg agtgttctct tggttaacttt tcttttatat    180
ttcctattct cctagcatgg atttaaaaaa gaaaaatatt cctatcctgc tcaactggtaa    240
ttaacatagg tttaaaatgg cttcaaatgt ggccctatag acggttaaaa ttgtacctta    300
tcttggcaaa acttcagagc accagtcagt gcatgcaagg tgccattttt tattgagatg    360
cttagaatgt ttctttctgt gcac                                384

```

What is claimed is:

1. An isolated nucleic acid encoding the amino acid 55 sequence of SEQ ID NO: 4.
2. An isolated oligonucleotide consisting of at least 18 contiguous nucleotides from an isolated nucleic acid encoding SEQ ID NO: 4, wherein said oligonucleotide includes a portion of the nucleic acid sequence that encodes a 60 glutamine at position 121 of SEQ NO: 4.
3. A method for detecting a predisposition to insulin resistance in an individual, the method comprising:

analyzing an individual for the presence of a genetic polymorphism in the genomic sequence of a human PC-1 allele, wherein said human PC-1 allele encodes a glutamine amino acid at position 121 of SEQ ID NO:4, and wherein the presence of said glutamine at position 121 of SEQ ID NO: 4 is indicative of a predisposition to insulin resistance.

* * * * *



US006228582B1

(12) **United States Patent**
Rodier et al.

(10) **Patent No.:** **US 6,228,582 B1**
 (45) **Date of Patent:** **May 8, 2001**

(54) **GENETIC POLYMORPHISMS WHICH ARE ASSOCIATED WITH AUTISM SPECTRUM DISORDERS**

(75) **Inventors:** Patricia M Rodier; Jennifer L. Ingram; Denise A. Figlewicz; Susan L. Hyman; Christopher J. Stodgell, all of Rochester, NY (US)

(73) **Assignee:** University of Rochester, Rochester, NY (US)

(*) **Notice:** Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) **Appl. No.:** 09/095,117

(22) **Filed:** Jun. 10, 1998

Related U.S. Application Data

(60) Provisional application No. 60/049,803, filed on Jun. 17, 1997.

(51) **Int. Cl.⁷** C12Q 1/68; C12P 19/34; C07H 21/04

(52) **U.S. Cl.** 435/6; 435/91.2; 536/23.1; 536/24.3

(58) **Field of Search** 435/6, 91.2; 536/23.1, 536/24.3

(56) **References Cited**

PUBLICATIONS

Jorde et al., *Am. J. Hum. Genet.* vol. 49, pp 932-938, 1991.*
 Goddard et al., "Mice With Targeted Disruption of *Hoxb-1* Fail to Form the Motor Nucleus of the VIIth Nerve," *Development*, 122:3217-3228 (1996).
 Studer et al., "Genetic Interactions Between *Hoxa1* and *Hoxb1* Reveal New Roles in Regulation of Early Hindbrain Patterning," *Development*, 125:1025-1036 (1998).
 Gavalas et al., "*Hoxa1* and *Hoxb1* Synergize in Patterning the Hindbrain, Cranial Nerves and Second Pharyngeal Arch," *Development*, 125:1123-1136 (1998).

Thompson et al., "An Evolutionary Conserved Element is Essential for Somite and Adjacent Mesenchymal Expression of The *Hoxa1* Gene," *Development Dynamics*, 211:97-108 (1998).

Spence et al., "Gene Mapping Studies with the Syndrome of Autism," *Behav. Genet.*, 15:1-13 (1985).

Smalley et al., "Autism and Genetics," *Arch. Gen. Psychiat.*, 45:953-961 (1988).

Cook et al., "Evidence of Linkage Between the Serotonin Transporter and Autistic Disorder," *Molec. Psychiat.*, 2:247-250 (1997).

Klauck et al., "Serotonin Transporter (5-HTT) Gene Variants Associated with Autism?," *Hum. Molec. Genet.*, 6(13):2233-2238 (1997).

Cook et al., "Linkage-Disequilibrium Mapping of Autistic Disorder, With 15q11-13 Markers," *Am. J. Hum. Genet.*, 62:1077-1083 (1998).

Acampora et al., "The Human HOX Gene Family," *Nucleic Acids Res.*, 17(24):10385-10402 (1989).

Hong et al., "Structure and Function of the *HOXA1* Human Homeobox Gene cDNA," *Gene*, 159:209-214 (1995).

Murphy et al., "Expression of the Mouse *Labial*-Like Homeobox-Containing Genes, *Hox 2.9* and *Hox 1.6*, During Segmentation of the Hindbrain," *Development*, 111:61-74 (1991).

* cited by examiner

Primary Examiner—W. Gary Jones

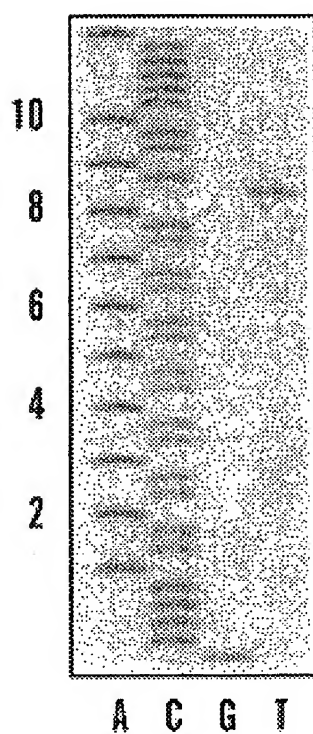
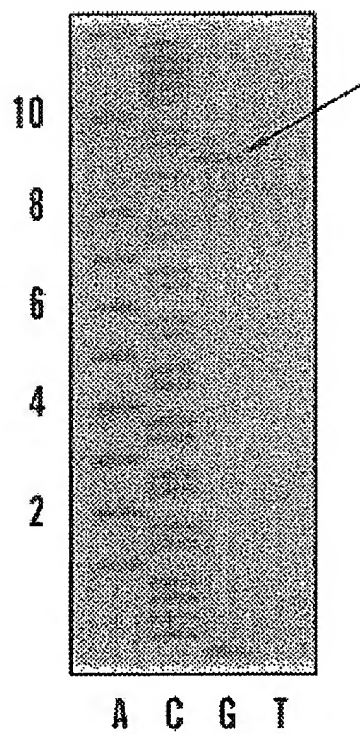
Assistant Examiner—Jehanne Souaya

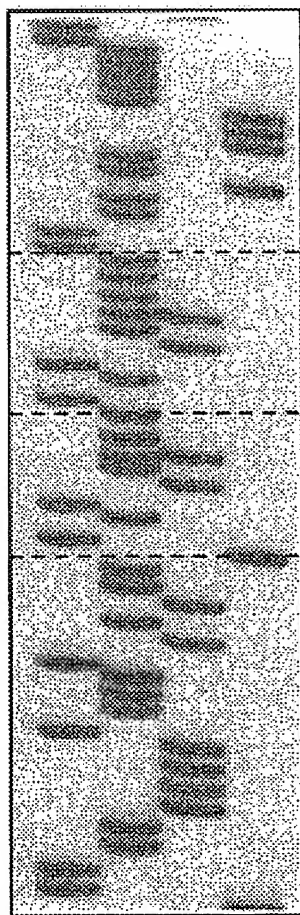
(74) *Attorney, Agent, or Firm*—Nixon Peabody LLP

(57) **ABSTRACT**

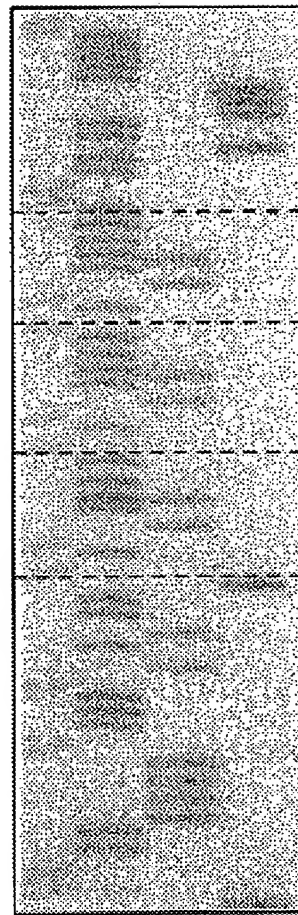
A method is provided for screening subjects for genetic markers associated with autism. The method involves isolating a biological sample from a mammal and then testing for the presence of a mutated gene or a product thereof which is associated with autism. Also disclosed are isolated nucleic acids encoding HoxA1 and HoxB1, both of which have a polymorphism that is associated with autism spectrum disorders.

29 Claims, 2 Drawing Sheets

**FIG. 1A****FIG. 1B**



A C G T

FIG. 2A

A C G T

FIG. 2B

GENETIC POLYMORPHISMS WHICH ARE ASSOCIATED WITH AUTISM SPECTRUM DISORDERS

The present application claims the benefit of U.S. Provisional Patent Application Ser. No. 60/049,803, filed Jun. 17, 1997.

The subject matter of this application was made with support from the United States Government under Grants No. RO1AA08666, RO1 NS 24287, RO1HD34295, RO1HD34969, and 2P30 ES01247 from the National Institutes of Health and Grant No. R824758 from the Environmental Protection Agency. The United States Government may retain certain rights.

FIELD OF THE INVENTION

The present invention relates to a method of screening subjects for genetic markers associated with autism. The invention further relates to isolated nucleic acids having polymorphisms associated with autism, the polypeptide products of those nucleic acids, and antibodies specific to the polypeptides produced by the mutated genes.

BACKGROUND OF THE INVENTION

Autism is a behaviorally defined syndrome characterized by impairment of social interaction, deficiency or abnormality of speech development, and limited activities and interest (American Psychiatric Association, 1994). The last category includes such abnormal behaviors as fascination with spinning objects, repetitive stereotypic movements, obsessive interests, and abnormal aversion to change in the environment. Symptoms are present by 30 months of age. The prevalence rate in recent Canadian studies using total ascertainment is over 1/1,000 (Bryson, S. E. et al., *J. Child Psychol. Psychiat.*, 29, 433 (1988)).

Attempts to identify the cause of the disease have been difficult, in part, because the symptoms do not suggest a brain region or system where injury would result in the diagnostic set of behaviors. Further, the nature of the behaviors included in the criteria preclude an animal model of the diagnostic symptoms and make it difficult to relate much of the experimental literature on brain injuries to the symptoms of autism.

Several quantitative changes have been observed in autistic brains at autopsy. An elevation of about 100 g in brain weight has been reported (Bauman, M. L. and Kemper, T. L., *Neurology* 35, 866 (1985)). While attempts to find anatomical changes in the cerebral cortex have been unsuccessful (Williams, R. S. et al., *Arch. Neurol.*, 37, 749 (1980); Coleman P. D., et al., *J. Autism Dev. Disord.*, 15, 245 (1985)), several brains have been found to have elevated neuron packing density in structures of the limbic system (Bauman, M. L. and Kemper, T. L., *Neurology* 35, 866 (1985)), including the amygdala, hippocampus, septal nuclei and mammillary body. Multiple cases in multiple labs have been found to have abnormalities of the cerebellum. A deficiency of Purkinje cell and granule cell number, as well as reduced cell counts in the deep nuclei of the cerebellum and neuron shrinkage in the inferior olive, have been reported (Bauman, M. L. and Kemper, T. L., *Neurology*, 35, 866 (1985); Bauman, M. L. and Kemper, T. L., *Neurology*, 36 (suppl. 1), 190 (1986); Bauman, M. L. and Kemper, T. L., *The Neurobiology of Autism*, Johns Hopkins University Press, 119 (1994); Ritvo, E. R. et al., *Am. J. Psychiat.*, 143, 862 (1986); Kemper, T. L. and Bauman M. L., *Neurobiology of Infantile Autism*, Elsevier Science Publishers, 43 (1992)).

Imaging studies have allowed examination of some anatomical characteristics in living autistic patients, providing larger samples than those available for histologic evaluation. In general, these confirm that the size of the brain in autistic individuals is not reduced and that most regions are also normal in size (Piven, J. et al., *Biol. Psychiat.*, 31, 491 (1992)). Reports of size reductions in the brainstem have been inconsistent (Gaffney, G. R. et al., *Biol. Psychiat.*, 24, 578 (1988); Hsu, M. et al., *Arch. Neurol.* 48, 1160 (1991)), but a new, larger study suggests that the midbrain, pons, and medulla are smaller in autistic cases than in controls (Hashimoto, T. et al., *J. Aut. Dev. Disord.*, 25, 1 (1995)). In light of the histological effects reported for the cerebellum, it is interesting that the one region repeatedly identified as abnormal in imaging studies is the neocerebellar vermis (lobules VI and VII; Gaffney, G. R. et al., *Am. J. Dis. Child.*, 141, 1330 (1987); Courchesne E., et al., *N. Engl. J. Med.*, 318, 1349 (1988); Hashimoto, T. et al., *J. Aut. Dev. Disord.*, 25, 1 (1995)). Not all comparisons have found a difference in neocerebellar size (Piven, J. et al., *Biol. Psychiat.*, 31, 491 (1992); Kleiman, M. D. et al., *Neurology*, 42, 753 (1992)), but a recent reevaluation of positive and negative studies (Courchesne, E. et al., *Neurology*, 44, 214 (1994)) indicates that a few autistic cases have hyperplasia of the neocerebellar vermis, while many have hypoplasia. Small samples of this heterogeneous population could explain disparate results regarding the size of the neocerebellum in autism. The proposal that the cerebellum in autistic cases can be either large or small is reasonable from an embryological standpoint, because injuries to the developing brain are sometimes followed by rebounds of neurogenesis (e.g., Andreoli, J. et al., *Am. J. Anat.* 137, 87 (1973); Bohn, M. C. and Lauder, J. M., *Dev. Neurosci.*, 1, 250 (1978); Bohn, M. C., *Neuroscience*, 5, 2003 (1980)), and it is possible that such rebounds could overshoot the normal cell number. Further, because increased cell density has been observed in the limbic system, the cerebellum is not the only brain region in which some form of overgrowth might account for the neuro-anatomy of autistic cases. It may well be that some autism-inducing injuries occur just prior to a period of rapid growth for the cerebellar lobules in question or the limbic system, leading to excess growth, while other injuries continue to be damaging during the period of rapid growth, leading to hypoplasia. However, the hypothesis that autism occurs with both hypoplastic and hyperplastic cerebella calls into question whether cerebellar anomalies play a major role in autistic symptoms.

A particularly instructive result has appeared in an MRI study on the cerebral cortex (Piven, J. et al., *Am. J. Psychiat.*, 14, 734 (1992)). Of a small sample of autistic cases, the majority showed gyral anomalies (e.g., patches of pachygyria). However, the abnormal areas were not located in the same regions from case to case. That is, while the functional symptoms were similar in all the subjects, the brain damage observed was not. The investigators argue convincingly that the cortical anomalies were not responsible for the functional abnormalities. This is a central problem in all attempts to screen for pathology in living patients or in autopsy cases. While abnormalities may be present, it is not necessarily true that they are related to the symptoms of autism.

To teratologists, the physical anomalies of a neonate, child, or adult can serve as a guide to when the embryo was injured. Years of research have amplified the details of that timetable for the nervous system (Rodier, P. M., *Dev. Med. Child Neurol.*, 22, 525 (1980); Bayer, S. A. et al., *Neurotoxicology*, 14, 83 (1993)). In the case of autism, lack

of specific information on the neuroanatomy associated with the disease has made it difficult to estimate the stage of development when the disorder arises. However, in 1993, Miller and Strömland reported a finding that conclusively identified the time of origin for some cases. They observed that the rate of autism was 33% in people exposed to thalidomide between the 20th and 24th days of gestation, and 0% in cases exposed at other times (Strömland, K. et al., *Devel. Med. Child. Neurol.*, 36, 351 (1994)). Their deduction regarding the time of injury was not based on neuroanatomy, which was not known in their living subjects. Instead, it was based on the external stigmata of the cases.

Because thousands of thalidomide-exposed offspring have been evaluated for somatic malformations, the array of injuries associated with the drug is well-known, and the time when each arises has been carefully defined (Miller, M. T., *Trans. Am. Ophthalmol. Soc.*, 89, 623 (1991)). Of five cases of thalidomide-induced autism, four had malformations of the ears, without limb malformation, and the fifth had malformation of the ears, forelimb, and hindlimb. Thalidomide is not teratogenic before the 20th day of gestation. Starting on day 20 exposure causes ear malformation and abnormalities of the thumb. Limb malformations (other than those of the thumb) first appear with exposure on the 25th day, with effects moving from the forelimb to the hindlimb as exposure occurs at later stages. After the 35th day, thalidomide produces no malformations. Thus, the cases with malformations restricted to the ear must have been exposed before day 25, and the one patient with multiple malformations can only be explained as a case of repeated injuries at several stages of development.

In fact, the idea that autism might arise very early in gestation was suggested long ago. Steg and Rapoport (*J. Aut. Child. Schiz.*, 5, 299 (1975)) noted the significant increase in minor physical anomalies among children with autism, and realized that they indicated an injury in the first trimester. Several studies of minor malformations have found ear effects to be the most common anomalies in autism (Walker, H. A., *J. Aut. Child. Schiz.*, 7, 165 (1977); Campbell, M. et al., *Am. J. Psychiat.*, 135, 573 (1978)), and the most recent study shows that they are not only the best discriminator between people with autism and normal controls, but also the only anomaly that discriminates autism from other developmental disabilities (Rodier, P. M. et al., *Teratology* 55, 319 (1997)). Ear anomalies are among the earliest of all minor physical malformations in their time of origin.

External malformations are not the only evidence which puts the time of injury in autism at the time of neural tube closure. The cranial nerve dysfunctions observed in the patients with autism secondary to thalidomide exposure—facial nerve palsy, Duane syndrome (lack of abducens innervation with reinnervation of the lateral rectus by the oculomotor nerve), abnormal lacrimation, gaze paresis, and hearing deficits (Strömland, K. et al., *Devel. Med. Child. Neurol.*, 36, 351 (1994))—suggest that the earliest-forming structures of the brain stem were damaged, and it is now known that these form during neural tube closure (Bayer, S. A. et al., *Neurotoxicology*, 14, 83 (1993)). Subsequent studies have shown that a human brain from a patient with autism has the same pattern of brain stem injury predicted by the thalidomide cases (Rodier, P. M. et al., *J. Comp. Neurol.*, 370, 247 (1996)). Perhaps even more importantly, the autopsied brain has a shortening of the brain stem in the region of the fifth rhombomere, and is missing two of the nuclei known to form from that embryological structure. The rhombomeres exist so briefly (Streeter, G. L., *Contr. Embryol. Carneg. Instn.*, 30, 213 (1948)) that the evidence

that one failed to form is conclusive in pinpointing the time of injury. Like the thalidomide cases, the autopsy case could have been injured only at the time of neural tube closure.

The effect of injury around neural tube closure has been tested experimentally, to see whether it can produce anatomical results like those suspected in the thalidomide cases and observed in human brain. Animals exposed during the critical period to valproic acid, a teratogen with effects similar to thalidomide, which has also been associated with autism (Christianson, A. L. et al., *Devel. Med. Child. Neurol.*, 36, 357 (1994); Williams, P. G. et al., *Dev. Med. Child. Neurol.*, 39, 632 (1997)) exhibit reductions in the number of cranial nerve motor neurons (Rodier, P. M. et al., *J. Comp. Neurol.*, 370, 247 (1996)). They are distinguished from controls by shortening of the hindbrain in the region which forms from the fifth rhombomere, just as the autopsied brain was (Rodier, P. M., et al., *Teratology* 55, 319 (1997)). Additional data suggests that the animal model has secondary changes in the cerebellum like those reported in some human cases of autism (Ingram, J. L. et al., *Teratology*, 53, 86 (1996)).

It has long been known that heritable factors play an important role in the etiology of autism. This was demonstrated by the original twin studies of Folstein and Rutter (*J. Child Psychol. Psychiat.*, 18, 297 (1977)) and the subsequent addition of more twin pairs to the sample has only increased the estimate of the proportion of cases suspected to have a genetic basis (e.g. Bailey, A. et al., *Psychol. Med.*, 25, 63 (1995); LeCouteur, A. et al., *J. Child Psychol. Psychiat.*, 37, 785 (1996)). Family studies of siblings (Smalley, S. L. et al., *Arch. Gen. Psychiat.*, 45, 953 (1988)) and parents (Landa, R. et al., *J. Speech Hear. Res.*, 34, 1339 (1991); Landa, R. et al., *Psych. Med.*, 22, 245 (1992)) also support the conclusion that an inherited risk is involved in many, perhaps all, cases of autism spectrum disorders. While the rate of autism is elevated in close relatives of cases, the rate of symptoms short of the diagnosis is increased much more. That is, individuals known to share genetic factors seem to vary in the degree to which symptoms are expressed. This non-Mendelian pattern (Jorde, L. B. et al., *Am. J. Hum. Genet.*, 49, 932 (1991)) suggests a complex disorder with major contributions from predisposing genetic factors, which interact with the overall genetic background and/or environmental insults to determine the phenotype.

The ability to identify the genetic factors that increase the risk for autism would be a breakthrough for genetic counseling for prevention of the disorder. In addition, it would allow the creation of genetically-engineered animals in which to study the environmental factors that interact with the inherited predispositions. Tests for genetic factors would also serve as biomarkers, valuable for diagnosis, and useful in research on all aspects of the autism spectrum. Unfortunately, neither linkage nor association studies have revealed any chromosomal regions strongly related to autism (e.g. Spence, M. A. et al., *Behav. Genet.*, 15, 1 (1985); Smalley, S. L. et al., *Arch. Gen. Psychiat.*, 45, 953 (1988); Cook, E. H. et al., *Molec. Psychiat.*, 2, 247 (1997); Klauck, S. M. et al., *Hum. Molec. Genet.*, 6, 2233 (1997); Cook, E. H. et al., *Am. J. Hum. Genet.*, 62, 1077 (1998)).

Furthermore, while there is no known medical treatment for autism, some success has been reported for early intervention with behavioral therapies. A biomarker would allow identification of the disease, now typically diagnosed between ages three and five, in infancy or prenatal life. Thus, there is an urgent need for a method of reliably identifying subjects with autism. In particular there is need for a blood test for polymorphisms causing autism spectrum disorders.

5

Families with affected members need to know whether they carry a mutation which could affect future pregnancies. Clinicians need a test as an aid in diagnosis, and researchers would use the test to classify subjects according to the etiology of their disease.

SUMMARY OF THE INVENTION

The present invention relates to a method for screening subjects for genetic markers associated with autism. A biological sample is isolated from a mammal and then tested for the presence of a mutated gene or a product thereof which is associated with autism.

Another aspect of the invention is an isolated nucleic acid encoding a HoxA1 allele having a polymorphism which is associated with autism spectrum disorders.

Yet another aspect of the invention is an isolated nucleic acid encoding a HoxB1 allele having a polymorphism which is associated with autism spectrum disorders.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows two different alleles of HoxA1 from a case of autism spectrum disorder. FIG. 1A shows the previously published sequence of wild-type HoxA1. FIG. 1B shows a previously unknown polymorphism having a single base substitution at position 218, where an A is changed to a G.

FIG. 2 shows a second polymorphism was identified in the first exon of HoxB1. The published sequence of wild-type HoxB1 (FIG. 2A) is compared to the previously unknown polymorphism in this paralog of HoxA1 (FIG. 2B). In this case, the anomaly is a nine-base insertion that adds a third repeat where two are normally present. The result is three extra amino acids, (serine-alanine-histidine). For each of the polymorphisms, it was possible to test for the

6

carried out on 30–40 subjects to be certain that the digestion results match the sequencing results, demonstrating that the digestion procedure detects the deviant sequence described and no other.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a method for screening subjects for genetic markers associated with autism. A biological sample is isolated from a mammal and then tested for the presence of a mutated gene or a product thereof which is associated with autism.

Polymorphisms in Hox genes are shown to be associated with autism spectrum disorders. The Hox genes are a family of genes that function in the patterning of body structures that develop along an anteroposterior axis, such as the limbs, skeleton, and nervous system; they are expressed during embryonic development at specific times in limited regions of the embryo. In the mouse, for example, Hox-a1 is expressed in rhombomeres 4 through 8 of the developing hindbrain on days 8 to 8.5 of gestation. The Hox genes control the pattern formation of the hindbrain. Similar abnormalities have been observed in the brains of autistic individuals (Rodier et al., *J. Comp. Neuro.* 370, 247 (1996), which is hereby incorporated by reference).

The DNA and amino acid sequences for HoxA-1 have previously been reported (Acampora, D. et al., *Nucleic Acids Res.*, 17, 10385 (1989); Hong, Y. et al., *Gene*, 159, 209 (1995) which are hereby incorporated by reference). Exon 1 stretches from base 1 to base 357. Exon 2 stretches from base 358 to the end (1008). The wildtype gene sequences for HoxA1 is provided in SEQ. ID. No. 1 as follows:

```

ATGGACAATG CAAGAATGAA CTCCTTCCTG GAATACCCCA TACTTAGCAG TGGCGACTCG      60
GGGACCTGCT CAGCCCGAGC CTACCCCTCG GACCATAGGA TTACAACTTT CCAGTCGTGC      120
GCGGTCAGCG CCAACAGTTG CGGCGGCGAC GACCGCTTCC TAGTGGGCAG GGGGTGTCAG      180
ATCGGTTGCG CCCACCACCA CCACCACCAC CACCATCACC ACCCCAGGCC GGCTACCTAC      240
CAGACTTCCG GGAACCTGGG GGTGTCTTAC TCCCACTCAA GTTGTGGTCC AAGCTATGGC      300
TCACAGAACT TCAGTGCGCC TTACAGCCCC TACGCGTTAA ATCAGGAAGC AGACGTAAGT      360
GGTGGGTACC CCCAGTGCGC TCCCGCTGTT TACTCTGGAA ATCTCTCATC TCCCATGGTC      420
CAGCATCACC ACCACCACCA GGGTTATGCT GGGGCGCGGG TGGGCTCGCC TCAATACATT      480
CACCCTCAT ATGGACAGGA GCACCAGAGC CTGGCCCTGG CTACGTATAA TAACTCCTTG      540
TCCCTCTCC ACGCCAGCCA CCAAGAAGCC TGTCGCTCCC CCGCATCGGA GACATCTTCT      600
CCAGCGCAGA CTTTGTACTG GATGAAAGTC AAAAGAAACC CTCCCAAAC AGGGAAGTT      660
GGAGAGTACG GCTACCTGGG TCAACCCAAC GCGGTGCGCA CCAACTTCAC TACCAAGCAG      720
CTCAGGAAC TGGAGAAGGA GTTCCACTTC AACAAGTACC TGACGCGCGC CCGCAGGGTG      780
GAGATCGCTG CATCCCTGCA GCTCAACGAG ACCCAAGTGA AGATCTGTTT CCAGAACCGC      840
CGAATGAAGC AAAAGAAAGC TGAGAAGGAG GGTCTCTTGC CCATCTCTCC GGCCACCCCG      900
CCAGGAACG ACGAGAAGGC CGAGGAATCC TCAGAGAAGT CCAGCTCTTC GCCCTGCGTT      960
CCTTCCCCGG GGTCTTCTAC CTCAGACACT CTGACTACCT CCCACTGA      1008

```

presence of the allele different from the known sequence by digesting PCR product with a restriction enzyme (Hph-I for HoxA1 and Msp-I for HoxB1). Sequencing reactions were

The nucleic acid molecule of SEQ. ID. No. 1 encodes a polypeptide having the amino acid sequence of SEQ. ID. No. 2, as follows:

M	D	N	A	R	M	N	S	F	L	E	Y	P	I	L	15
S	S	G	D	S	G	T	C	S	A	R	A	Y	P	S	30
D	H	R	I	T	T	F	Q	S	C	A	V	S	A	N	45
S	C	G	G	D	D	R	F	L	V	G	R	G	V	Q	60
I	G	S	P	H	H	H	H	H	H	H	H	H	H	P	75
Q	P	A	T	Y	Q	T	S	G	N	L	G	V	S	Y	90
S	H	S	S	C	G	P	S	Y	G	S	Q	N	F	S	105
A	P	Y	S	P	Y	A	L	N	Q	E	A	D	V	S	120
G	G	Y	P	Q	C	A	P	A	V	Y	S	G	N	L	135
S	S	p	M	V	Q	H	H	H	H	H	Q	G	Y	A	150
G	G	A	V	G	S	P	Q	Y	I	H	H	S	Y	G	165
Q	E	H	Q	S	L	A	L	A	T	Y	N	N	S	L	180
S	P	L	H	A	S	H	Q	E	A	C	R	S	P	A	195
S	E	T	S	S	P	A	Q	T	F	D	W	M	K	V	210
K	R	N	P	P	K	T	G	K	V	G	E	Y	G	Y	225
L	G	Q	P	N	A	V	R	T	N	F	T	T	K	Q	240
L	T	E	L	E	K	E	F	H	F	N	K	Y	L	T	255
R	A	R	R	V	E	I	A	A	S	L	Q	L	N	E	270
T	Q	V	K	I	W	F	Q	N	R	R	M	K	Q	K	285
K	R	E	K	E	G	L	L	P	I	S	P	A	T	P	300
P	G	N	D	E	K	A	E	E	S	S	E	K	S	S	315
S	S	P	C	V	P	S	P	G	S	S	T	S	D	T	330
L	T	T	S	H											335

25

A polymorphism in the HocA1 gene has been isolated and sequenced. This polymorphism is associated with autism spectrum disorders. A single base substitution is located at position 218 (underlined) of SEQ. ID. No. 3, where an A is changed to a G, as follows:

The single base substitution at position 218 results in the replacement of histidine with arginine (underlined). The resulting protein has the amino acid sequence (SEQ. ID. No. 4) as follows:

ATGGACAATG	CAAGAATGAA	CTCCTTCCTG	GAATACCCCA	TACTTAGCAG	TGGCGACTCG	60
GGGACCTGCT	CAGCCCAGAG	CTACCCCTCG	GACCATAGGA	TTACAACTTT	CCAGTCGTGC	120
GCGGTCAGCG	CCAACAGTTG	CGGCGGCGAC	GACCGCTTCC	TAGTGGGCAG	GGGGGTGCAG	180
ATCGGTCGCG	CCCACCACCA	CCACCACCAC	CACCATCGCC	ACCCCGAGCC	GGCTACCTAC	240
CAGACTTCGG	GGAACCTGGG	GGTGTCTTAC	TCCCACTCAA	GTTGTGGTCC	AAGCTATGGC	300
TCACAGAACT	TCAGTGCGCC	TTACAGCCCC	TACGCGTTAA	ATCAGGAAGC	AGACGTAAGT	360
GGTGGGTACC	CCCAGTGCAG	TCCCGCTGTT	TACTCTGGAA	ATCTCTCATC	TCCCATGGTC	420
CAGCATCACC	ACCACCACCA	GGGTTATGCT	GGGGGCGCGG	TGGGCTCGCC	TCAATACATT	480
CACCACTCAT	ATGGACAGGA	GCACCAGAGC	CTGGCCCTGG	CTACGTATAA	TAATCCTTTG	540
TCCCTCTCTC	ACGCCAGCCA	CCAAGAAGCC	TGTCGCTCCC	CCGCATCGGA	GACATCTTCT	600
CCAGCGCAGA	CTTTTGACTG	GATGAAAGTC	AAAAGAAACC	CTCCCAAAAC	AGGGAAAGTT	660
GGAGAGTACG	GCTACCTGGG	TCAACCCAAC	GCGGTGCGCA	CCAACTTCAC	TACCAAGCAG	720
CTCACGGAAC	TGGAGAAGGA	GTTCCACTTC	AACAAGTACC	TGACGCGCGC	CCGCAGGGTG	780
GAGATCGCTG	CATCCCTGCA	GCTCAACGAG	ACCCAAGTGA	AGATCTGGTT	CCAGAACCGC	840
CGAATGAAGC	AAAAGAAACG	TGAGAAGGAG	GGTCTCTTGC	CCATCTCTCC	GGCCACCCCG	900
CCAGGAACG	ACGAGAAGGC	CGAGGAATCC	TCAGAGAAGT	CCAGCTCTTC	GCCCTGCGTT	960
CCTTCCCCGG	GGTCTTCTAC	CTCAGACACT	CTGACTACCT	CCCACTGA		1008

60

In addition to the polymorphism in HoxA1, a polymorphism associated with autism spectrum disorders has been isolated and sequenced from the HoxB1 gene. The Hoxb1 gene has not been studied as comprehensively as Hoxa1 in transgenic knockouts, but is expressed at the same stage (Murphy, P et al., *Development*, 111, 61 (1991), which is hereby incorporated by reference). Its null mutation produces similar malformations, including severe diminution of the facial nucleus (Goddard, J. M. et al., *Development*, 122, 3217 (1996), which is hereby incorporated by reference).

The similarity of expression and function of these two genes is due to the fact that they were originally a single gene in invertebrates (Ruddle, F. H. et al., *Annu. Rev. Genet.*, 28, 423 (1993), which is hereby incorporated by reference). In mammals, the two appear on separate chromosomes (human 7 and 17), but the sequence of each of the mammalian genes is similar to the others, and similar to the original single gene from which the two mammalian loci arose. The sequence of the wildtype *hoxB1* gene (SEQ. ID. No. 5) follows:

TGACGCGATGG	ACTATAATAG	GATGAACTCC	TTCTTAGAGT	ACCCACTCTG	TAACCGGSGA	60
CCCAGCGCCT	ACAGCGCCCA	CAGCGCCCA	ACCTCCTTTC	CCCCAAGCTC	GGCTCAGGCG	120
GTTGACAGCT	ATGCAAGCGA	GGGCCGTAC	GGTGGGGGGC	TGTCCAGCCC	TGCGTTTCAG	180
CAGAACTCCG	GCTATCCCGC	CCAGCAGCCG	CCTTCGACCC	TGGGGGTGCC	CTTCCCCAGC	240
TCCGCGCCCT	CGGGGTATGC	TCCTGCCGCC	TGCAGCCCCA	GCTACGGGCC	TTCTCAGTAC	300
TACCTCTCTG	GTCAATCAGA	AGGAGACGGA	GGCTATTTC	ATCCCTCGAG	CTACGGGGCC	360
CAGCTAGGGG	GCTTGTCCGA	TGGCTACGGA	GCAGGTGGAG	CCGGTCCGGG	GCCATATCCT	420
CCGCAGCATC	CCCCTTATGG	GAACGAGCAG	ACCGCGAGCT	TTGCACCGGC	CTATGCTGAT	480
CTCCTCTCCG	AGGACAAGGA	AACACCTGCG	CCTTCAGAAG	CTAACACCCC	CACGGCCCGG	540
ACCTTCGACT	GGATGAAGGT	TAAGAGAAAC	CCACCCAAGA	CAGCGAAGGT	GTCAGAGCCA	600
GGCCTGGGCT	CGCCAGTGG	CCTCCGCACC	AACTTCACCA	CAAGGCAGCT	GACAGAAGCTG	660
GAAAGGAGT	TCCATTTCAA	CAAGTACCTG	AGCCGGGCCC	GGAGGGTSGA	GATTGCCGCC	720
ACCTTGAGAG	TCAATGAAAC	ACAGGTCAAG	ATTTGGTTCC	AGAACCAGAC	AATGAAGCAG	780
AAGAAGCGCG	AGCGAGAGGG	AGGTCGGGTC	CCCCCAGCCC	CACCAGGCTG	CCCCAAGGAG	840
GCAGCTGGAG	ATGCCTCAGA	CCAGTCGACA	TGCACCTCCC	CGGAAGCCTC	ACCCAGCTCT	900
GTCACCTCCT	GAACTGAACC	TAGCCACCAA	TGGGGCTTCC	AGGCACCTGA	GCGCCCCAGT	960
CCAGCCCTAT	CCCAGGCTCT	CCCAACCCAG	GCCTGGCTTC	ACTGCCTGGG	ATCTCTAGGC	1020
T						1021

The protein encoded by nucleotides 7 to 909 of the wild-type HoxB1 gene (SEQ. ID. No. 6) is as follows:

M	D	Y	N	R	M	N	S	F	L	E	Y	P	L	C	15
N	R	G	P	S	A	Y	S	A	H	S	A	P	T	S	30
F	P	P	S	S	A	Q	A	V	D	S	Y	A	S	E	45
G	R	Y	D	G	G	L	S	S	P	A	F	Q	Q	N	60
S	G	Y	P	A	Q	Q	P	P	S	T	L	G	V	P	75
E	P	S	S	A	P	S	G	Y	A	P	A	A	C	S	90
P	S	Y	G	P	S	Q	Y	Y	P	L	G	Q	S	E	105
G	D	G	G	Y	F	H	P	S	S	Y	G	A	Q	L	120
G	G	L	S	D	G	Y	G	A	G	G	A	G	P	G	135
P	Y	P	P	Q	H	P	P	Y	G	N	E	Q	T	A	150
S	F	A	P	A	Y	A	D	L	L	S	E	D	K	E	165
T	P	C	P	S	E	P	N	T	P	T	A	R	T	E	180
D	W	M	K	V	K	R	N	P	P	K	T	A	K	V	195
S	E	P	G	L	G	S	P	S	G	L	R	T	N	F	210
T	T	R	Q	L	T	E	L	E	K	E	F	H	F	N	225
K	Y	L	S	R	A	R	R	V	E	I	A	A	T	L	240
E	L	N	E	T	Q	V	K	I	W	F	Q	N	R	R	255
M	K	Q	K	K	R	E	R	E	G	G	R	V	P	F	270
A	P	P	G	C	P	K	E	A	A	G	D	A	S	D	285
Q	S	T	C	T	S	P	E	A	S	P	S	S	V	T	300
S															301

insertion is such that the amino acid sequence also changes. The normal sequence reads . . . serine-alanine-histidine-

As with the HoxA1 gene, polymorphisms associated with autism spectrum disorders were found with HoxB1. The HoxB1 mutation occurs after base 88 (C) with the insertion of nine nucleotides (ACAGCGCCC). The location of this

serine-alanine-proline. The mutant sequence has an extra serine-alanine-histidine-sequence and then the sequence resumes normally. The insertion and altered amino acid sequence are underlined below. A mutated form of HoxB1 (SEQ. ID. No. 7) is depicted as follows:

TGACGCATGG	ACTATAATAG	GATGAACTCC	TTCTTAGAGT	ACCCACTCTG	TAACCGGGGA	60
CCCAGCGCCT	ACAGCGCCCA	CAGCGCCCA	<u>AGCGCCCA</u>	CCTCCTTTCC	CCCAAGCTCG	120
GCTCAGGCGG	TTGACAGCTA	TGCAAGCGAG	GGCCGCTACG	GTGGGGGGCT	GTCCAGCCCT	180
GCGTTTCAGC	AGAACTCCGG	CTATCCCGCC	CAGCAGCCGC	CTTCGACCCT	GGGGGTGCC	240
TTCCCCAGCT	CGCGCCCTC	GGGTATGCT	CCTGCCGCCT	GCAGCCCCAG	CTACGGGCCT	300
TCTCAGTACT	ACCCTCTGGG	TCAATCAGAA	GGAGACGGAG	GCTATTTTCA	TCCCTCGAGC	360
TACGGGGCCC	AGCTAGGGGG	CTTGTCGGAT	GGCTACGGAG	CAGGTGGAGC	CGGTCCGGGG	420
CCATATCCTC	CGCAGCATCC	CCCTTATGGG	AACGAGCAGA	CCGCGAGCTT	TGCACCGGCC	480
TATGCTGATC	TCCTCTCCGA	GGACAAGGAA	ACACCCCTGCC	CTTCAGAAAC	TAACACCCCC	540
ACGGCCCGGA	CCTTCGACTG	GATGAAGGTT	AAGAGAAACC	CACCCAAGAC	AGCGAAGGTG	600
TCAGAGCCAG	GCCTGGGCTC	GCCAGTGGC	CTCCGCACCA	ACTTCACCAC	AAGGCAGCTG	660
ACAGAACTGG	AAAAGGAGTT	CCATTTC AAC	AAGTACCTGA	GCCGGGCCCG	GAGGGTGGAG	720
ATTGCCGCCA	CCCTGGAGCT	CAATGAAACA	CAGGTCAAGA	TTTGGTTCCA	GAACCGACGA	780
ATGAAGCAGA	AGAAGCGCGA	GCGAGAGGGA	GGTCGGGTCC	CCCCAGCCCC	ACCAGGCTGC	840
CCCAAGGAGG	CAGCTGGAGA	TGCTCAGAC	CAGTCGACAT	GCACCTCCCC	GGAAGCCTCA	900
CCCAGCTCTG	TCACCTCCTG	AACTGAACCT	AGCCACCAAT	GGGGCTTCCA	GGCACTGGAG	960
CGCCCCAGTC	CAGCCCTATC	CCAGGCTCTC	CCAACCCAGG	CCTGGCTTCA	CTGCCTGGGA	1020
TCTCTAGGCT						1030

The protein encoded by SEQ. ID. No. 8 is as follows:

M	D	Y	N	R	M	N	S	F	L	E	Y	P	L	C	15
N	R	G	P	S	A	Y	S	A	H	S	A	H	S	A	30
P	T	S	F	P	P	S	S	A	Q	A	V	D	S	Y	45
A	S	E	G	R	Y	G	G	G	L	S	S	P	A	F	60
Q	Q	N	S	G	Y	P	A	Q	Q	P	P	S	T	L	75
G	V	P	F	P	S	S	A	P	S	G	Y	A	P	A	90
A	C	S	P	S	Y	G	P	S	Q	Y	Y	P	L	G	105
Q	S	E	G	D	G	G	Y	F	H	P	S	S	Y	G	120
A	Q	L	G	G	L	S	D	G	Y	G	A	G	G	A	135
G	P	G	P	Y	P	P	Q	H	P	P	Y	G	N	E	150
Q	T	A	S	F	A	P	A	Y	A	D	L	L	S	E	165
D	K	E	T	P	C	P	S	E	P	N	T	P	T	A	180
R	T	F	D	W	M	K	V	K	R	N	P	P	K	T	195
A	K	V	S	E	P	G	L	G	S	P	S	G	L	R	210
T	N	F	T	T	R	Q	L	T	E	L	E	K	E	F	225
H	F	N	K	Y	L	S	R	A	R	R	V	E	I	A	240
A	T	L	E	L	N	E	T	Q	V	K	I	W	F	Q	255
N	R	R	M	K	Q	K	K	R	E	R	E	G	G	R	270
V	P	P	A	P	P	G	C	P	K	E	A	A	G	D	285
A	S	D	Q	S	T	C	T	S	P	E	A	S	P	S	300
S	V	T	S												304

Genes which have been duplicated and then maintained similar functions over the course of evolution are called "paralogs." A third paralog derived from the same invertebrate gene is known as HoxD1. This gene has not yet been studied in knockouts, but is known to have evolved to be expressed in somewhat different embryonic tissues (mesoderm vs. ectoderm) in the hindbrain region at the same stage of development as Hoxa1 and Hoxb1. Thus preferred hox genes include HoxA1, HoxB1, and HoxD1.

Biological samples suitable for testing include blood, saliva, amniotic fluid, and tissue. The most preferred biological sample is blood. However, any biological sample from which genetic material or the products of the marker genes can be isolated is suitable.

Because the Hox genes are highly conserved among species, the present invention is applicable for screening for autism related polymorphisms in mammals. The screening method can be utilized to identify animals carrying defects in genes like those which give rise to autism in humans in order to study the progression of the disease and test treatments. However, the preferred mammal to be screened is humans. In particular, the biological samples are isolated from developmentally disabled children or adults in order to determine whether they carry the marker associated with autism to assist in diagnosing the disease. Similarly, the parents or relatives of disabled children may be screened to determine whether they are carriers of the mutated gene. Samples may also be tested from children including infants to identify those children who have genetic markers associated with autism in order to provide them with early behavior training.

As discussed more fully in the examples, polymorphisms in the HoxA1 gene are associated with autism spectrum disorders. In addition to HoxA1, the HoxB1 and HoxD1 genes are also involved in the same stages of early brain development. Hoxb1 and Hoxd1 are related developmental genes which are expressed at the same time and in approximately the same region of the embryo as Hoxa1. The Hox genes are closely related and may perform similar functions in development. Evolutionarily the various Hox genes were probably derived from a common ancestral gene. Thus, the preferred genes to be screened include Hoxa1, Hoxb1, and Hoxd1.

The mutation in the mutated gene may be a single base substitution mutation resulting in an amino acid substitution, a single base substitution mutation resulting in a translational stop, an insertion mutation, a deletion mutation, or a gene rearrangement. As demonstrated from the identified polymorphisms in HoxA1 and HoxB1, polymorphisms which disrupt the gene or result in an altered peptide are associated with autism spectrum disorders.

The mutation may be located in an intron, an exon of the gene, or a promotor or other regulatory region which affects the expression of the gene.

Methods for screening for mutated nucleic acids include direct sequencing of nucleic acids, single strand polymorphism assay, ligase chain reaction, enzymatic cleavage, and southern hybridization.

Screening for mutated nucleic acids can be accomplished by direct sequencing of nucleic acids. In fact, putative mutants identified by other methods may be sequenced to determine the exact nature of the mutation. Nucleic acid sequences can be determined through a number of different techniques which are well known to those skilled in the art. In order to sequence the nucleic acid, sufficient copies of the material must first be amplified.

Amplification of a selected, or target, nucleic acid sequence may be carried out by any suitable means. (See generally Kwok, D. and Kwok, T., *Am Biotechnol Lab*, 8, 14 (1990), which is hereby incorporated by reference.) Examples of suitable amplification techniques include, but are not limited to, polymerase chain reaction, ligase chain reaction (see Barany, *Proc Natl Acad Sci USA* 88, 189 (1991), which is hereby incorporated by reference), strand displacement amplification (see generally Walker, G. et al., *Nucleic Acids Res.* 20, 1691 (1992); Walker, G. et al., *Proc Natl Acad Sci USA* 89, 392 (1992), which are hereby incorporated by reference), transcription-based amplification (see Kwok, D. et al., *Proc Natl Acad Sci USA*, 86, 1173 (1989), which is hereby incorporated by reference), self-sustained sequence replication (or "3SR") (see Guatelli, J. et al., *Proc Natl Acad Sci USA*, 87, 1874 (1990), which is hereby incorporated by reference), the Q β replicase system (see Lizardi, P. et al., *Biotechnology*, 6, 1197 (1988), which is hereby incorporated by reference), nucleic acid sequence-based amplification (or "NASBA") (see Lewis, R., *Genetic*

Engineering News, 12(9), 1 (1992), which is hereby incorporated by reference), the repair chain reaction (or "RCR") (see Lewis, R., *Genetic Engineering News*, 12(9), 1 (1992), which is hereby incorporated by reference), and boomerang DNA amplification (or "BDA") (see Lewis, R., *Genetic Engineering News*, 12(9), 1 (1992), which is hereby incorporated by reference). Polymerase chain reaction is currently preferred.

In general, DNA amplification techniques such as the foregoing involve the use of a probe, a pair of probes, or two pairs of probes which specifically bind to DNA encoding the gene of interest, but do not bind to DNA which does not encode the gene, under the same hybridization conditions, and which serve as the primer or primers for the amplification of the gene of interest or a portion thereof in the amplification reaction.

Nucleic acid sequencing can be performed by chemical or enzymatic methods. The enzymatic method relies on the ability of DNA polymerase to extend a primer, hybridized to the template to be sequenced, until a chain-terminating nucleotide is incorporated. The most common methods utilize dideoxynucleotides. Primers may be labelled with radioactive or fluorescent labels. Various DNA polymerases are available including Klenow fragment, AMV reverse transcriptase, *Thermus aquaticus* DNA polymerase, and modified T7 polymerase.

Although DNA sequencing is clearly the most sensitive and informative method, it is too cumbersome for routine use in searching for polymorphisms, especially when the DNA segment of interest is large. Several other methods are available for a rapid search for changes in autism associated genes.

Recently, single strand polymorphism assay ("SSPA") analysis and the closely related heteroduplex analysis methods have come into use as effective methods for screening for single-base polymorphisms (Orita, M. et al., *Proc Natl Acad Sci USA*, 86, 2766 (1989), which is hereby incorporated by reference). In these methods, the mobility of PCR-amplified test DNA from clinical specimens is compared with the mobility of DNA amplified from normal sources by direct electrophoresis of samples in adjacent lanes of native polyacrylamide or other types of matrix gels. Single-base changes often alter the secondary structure of the molecule sufficiently to cause slight mobility differences between the normal and mutant PCR products after prolonged electrophoresis.

Ligase chain reaction is yet another recently developed method of screening for mutated nucleic acids. Ligase chain reaction (LCR) is also carried out in accordance with known techniques. LCR is especially useful to amplify, and thereby detect, single nucleotide differences between two DNA samples. In general, the reaction is carried out with two pairs of oligonucleotide probes: one pair binds to one strand of the sequence to be detected; the other pair binds to the other strand of the sequence to be detected. The reaction is carried out by, first, denaturing (e.g., separating) the strands of the sequence to be detected, then reacting the strands with the two pairs of oligonucleotide probes in the presence of a heat stable ligase so that each pair of oligonucleotide probes hybridize to target DNA and, if there is perfect complementarity at their junction, adjacent probes are ligated together. The hybridized molecules are then separated under denaturation conditions. The process is cyclically repeated until the sequence has been amplified to the desired degree. Detection may then be carried out in a manner like that described above with respect to PCR.

Southern hybridization is also an effective method of identifying differences in sequences. Hybridization conditions, such as salt concentration and temperature can be adjusted for the sequence to be screened. Southern blotting and hybridizations protocols are described in *Current Protocols in Molecular Biology* (Greene Publishing Associates and Wiley-Interscience), pages 2.9.1-2.9.10. Probes can be labelled for hybridization with random oligomers (primarily 9-mers) and the Klenow fragment of DNA polymerase. Very high specific activity probe can be obtained using commercially available kits such as the Ready-To-Go DNA Labelling Beads (Pharmacia Biotech), following the manufacturer's protocol. Briefly, 25 ng of DNA (probe) is labelled with ³²P-dCTP in a 15 minute incubation at 37° C. Labelled probe is then purified over a ChromaSpin (Clontech) nucleic acid purification column. Possible competition of probes having high repeat sequence content, and stringency of hybridization and washdown will be determined individually for each probe used. Alternatively, fragments of a candidate gene may be generated by PCR, the specificity may be verified using a rodent-human somatic cell hybrid panel, and subcloning the fragment. This allows for a large prep for sequencing and use as a probe. Once a given gene fragment has been characterized, small probe preps can be done by gel- or column-purifying the PCR product.

These mismatch detection protocols use samples generated by PCR and thus require use of very little genomic template. All of these methods can provide very good clues regarding the location of the sequence change which leads to the appearance of anomalous bands, hence facilitating subsequent cloning and sequencing strategies.

Methods of screening for mutated nucleic acids can be carried out using either deoxyribonucleic acids ("DNA") or messenger ribonucleic acids ("mRNA") isolated from the biological sample. During periods when the gene is expressed, mRNA may be abundant and more readily detected. However, these genes are temporally controlled and, at most stages of development, the preferred material for screening is DNA.

Alternatively, the detection of a mutated gene associated with autism can be carried out by collecting a biological sample and testing for the presence or form of the protein produced by the gene. The mutation in the gene may result in the production of a mutated form of the peptide or the lack of production of the gene product. In this embodiment, the determination of the presence of the polymorphic form of the protein can be carried out, for example, by isoelectric focusing, protein sizing, or immunoassay. In an immunoassay, an antibody that selectively binds to the mutated protein can be utilized (for example, an antibody that selectively binds to the mutated form of HoxA1 encoded protein). Such methods for isoelectric focusing and immunoassay are well known in the art, and are discussed in further detail below.

Changes in the size or charge of the polypeptide can be identified by isoelectric focusing or protein sizing techniques. Changes resulting in amino acid substitutions, where the substituted amino acid has a different charge than the original amino acid, can be detected by isoelectric focusing. Isoelectric focusing of the polypeptide through a gel having an ampholine gradient at high voltages separates proteins by their pI. The pH gradient gel can be compared to a simultaneously run gel containing the wild-type protein. Protein sizing techniques such as protein electrophoresis and sizing chromatography can also be used to detect changes in the size of the product.

As an alternative to isoelectric focusing or protein sizing, the step of determining the presence of the mutated polypeptides in a sample may be carried out by an antibody assay with an antibody which selectively binds to the mutated polypeptides (i.e., an antibody which binds to the mutated polypeptides but exhibits essentially no binding to the wild-type polypeptide without the polymorphism in the same binding conditions).

Antibodies used to bind selectively the products of the mutated genes can be produced by any suitable technique. For example, monoclonal antibodies may be produced in a hybridoma cell line according to the techniques of Kohler and Milstein, *Nature*, 265, 495 (1975), which is hereby incorporated by reference. A hybridoma is an immortalized cell line which is capable of secreting a specific monoclonal antibody. The mutated products of genes which are associated with autism may be obtained from a human patient, purified, and used as the immunogen for the production of monoclonal or polyclonal antibodies. Purified polypeptides may be produced by recombinant means to express a biologically active isoform, or even an immunogenic fragment thereof may be used as an immunogen. Monoclonal Fab fragments may be produced in *Escherichia coli* from the known sequences by recombinant techniques known to those skilled in the art. (See, e.g., Huse, W., *Science* 246, 1275 (1989), which is hereby incorporated by reference) (recombinant Fab techniques).

The term "antibodies" as used herein refers to all types of immunoglobulin, including IgG, IgM, IgA, IgD, and IgE. The antibodies may be monoclonal or polyclonal and may be of any species of origin, including (for example) mouse, rat, rabbit, horse, or human, or may be chimeric antibodies, and include antibody fragments such as, for example, Fab, F(ab')₂, and Fv fragments, and the corresponding fragments obtained from antibodies other than IgG.

Antibody assays may, in general, be homogeneous assays or heterogeneous. In a homogeneous assay the immunological reaction usually involves the specific antibody, a labeled analyte, and the sample of interest. The signal arising from the label is modified, directly or indirectly, upon the binding of the antibody to the labeled analyte. Both the immunological reaction and detection of the extent thereof are carried out in a homogeneous solution. Immunochemical labels which may be employed include free radicals, radioisotopes, fluorescent dyes, enzymes, bacteriophages, coenzymes, and so forth.

In a heterogeneous assay approach, the reagents are usually the specimen, the antibody of the invention and means for producing a detectable signal. Similar specimens as described above may be used. The antibody is generally immobilized on a support, such as a bead, plate, or slide, and contacted with the specimen suspected of containing the antigen in a liquid phase. The support is then separated from the liquid phase and either the support phase or the liquid phase is examined for a detectable signal employing means for producing such signal. The signal is related to the presence of the analyte in the specimen. Means for producing a detectable signal include the use of radioactive labels, fluorescent labels, enzyme labels, and so forth. For example, if the antigen to be detected contains a second binding site, an antibody which binds to that site can be conjugated to a detectable group and added to the liquid phase reaction solution before the separation step. The presence of the detectable group on the solid support indicates the presence of the antigen in the test sample. Examples of suitable immunoassays are the radioimmunoassay, immunofluorescence methods, enzyme-linked immunoassays, and the like.

Those skilled in the art will be familiar with numerous specific immunoassay formats and variations thereof which may be useful for carrying out the method disclosed herein. See U.S. Pat. Nos. 4,727,022, 4,659,678, 4,376,110, 4,275, 149, 4,233,402, and 4,230,767.

Antibodies which selectively bind a polymorphic DLST isoform may be conjugated to a solid support suitable for a diagnostic assay (e.g., beads, plates, slides or wells formed from materials such as latex or polystyrene) in accordance with known techniques, such as precipitation. Antibodies which bind a polymorphic DLST isoform may likewise be conjugated to detectable groups such as radiolabels (e.g., ³⁵S, ¹²⁵I, ¹³¹I), enzyme labels (e.g., horseradish peroxidase, alkaline phosphatase), and fluorescent labels (e.g., fluorescein) in accordance with known techniques.

The invention further provides an isolated nucleic acid molecule which encodes a HoxA1 gene having a single base substitution at nucleotide 218 in SEQ. ID. No. 1. In another embodiment, the invention provides an isolated nucleic acid molecule which encodes a HoxB1 gene having an insertion between positions nucleotides 88 and 89 in SEQ. ID. No. 5. In addition, the invention provides fragments of the HoxA1 and HoxB1 genes having the polymorphism, where the fragment has at least 15 nucleotides and encompasses the polymorphism, i.e., the single base substitution. Fragments longer than 15 nucleotides can be used to probe for nucleic acid molecules containing the polymorphism. Longer fragments may be used at higher stringency conditions.

The invention also provides isolated polypeptides that are encoded by the genes having the polymorphisms. Either the whole protein or fragments thereof may be used to induce the production of antibodies specific to the portion of the protein which is effected by the polymorphism. Such antibodies may then be used to detect the presence of a polymorphism. Preferred antibodies bind specifically to the protein or polypeptide effected by the polymorphism but with less affinity to the wild-type Hox protein.

In one embodiment, the antibody is a monoclonal antibody. For use in an immunoassay, the antibody can be bound to a solid support or bound to a detectable label.

EXAMPLES

Example 1

Collection of Blood Samples from Autistic Individuals

Blood was collected from patients with autism and their immediate family members in order to determine whether any polymorphisms in HoxA1 are present among this population. All blood samples were procured following written consent by the patients or their guardians. Among the samples collected were those of the members of a family of four in which one child has autism and the other has Asperger's syndrome; both children have malformed ears. The first son is retarded and the second has normal intelligence. The parents have no obvious symptoms. DNA was extracted from the blood by phenolchloroform extraction following isolation and lysis of the white blood cells. Control DNA was also used for these excrements; this DNA was obtained from neurologically normal donors.

The 20 cc blood samples were left for three-four days at room temperature to allow continued proliferation of white blood cells. White cells were pelleted, followed by isolation of the nuclei. The nuclei were then incubated overnight at 37° C. in a lysis buffer consisting of EDTA, TNE-SDS, and proteinase K. Protein contaminants were extracted by additions of buffered phenol followed by chloroform, then DNA was precipitated by the addition of ice-cold ethanol. The

19

DNA was resuspended in TE buffer for storage at 4° C. Extraction of genomic DNA from fixed tissue was carried out using the protocol of Volkenandt et al., *Methods in Molecular Biology*, 15, 81, Humana Press, (1993), which is hereby incorporated by reference).

Example 2

Sequencing the HoxA1 Gene

The HoxA1 gene was amplified by PCR from DNA samples to provide sufficient material for sequencing. Two sets of oligonucleotide primers were selected after examination of the human HoxA1 nucleic acid sequence and comparison of the sequence to those of human and mouse Hox genes. The first set was designed to amplify residues 10–647, the second to amplify from residue 656 to the stop codon at residue 1008, exons 1 and 2 of HoxA1, respectively. The primers were used in polymerase chain reaction to amplify the target gene in several control blood samples, in order to determine the appropriate PCR conditions. Both exons were amplified by 94° C. denaturation for 1 min, 62° C., annealing for 30 sec, and 72° C. extension for 2 min, for 35 cycles. The products were visualized with ethidium bromide staining on a 1–2% agarose gel. PhiX174 RF DNA/Hae III fragments (Gibco) were used as a molecular weight marker. The products were tested for chromosome origin by using human-rodent monochromosomal somatic cell hybrids. Both exons amplified by the HoxA1 primers amplified the hybrid containing human chromosome 7 and do not amplify from any other hybrids. Establishing that the product amplified by the primers is from the correct chromosome rules out the possibility that pseudogenes with the same sequence occur at other sites or that the amplified product is another homologous homeobox gene. It verifies that the PCR product represents only the targeted gene.

The polymerase chain reaction (PCR) was performed with various samples of control DNA in order to determine the appropriate conditions. Once the optimal conditions were ascertained, the gene was amplified from the patient samples.

Following PCR, an aliquot of the product was used for DNA sequencing using the Sequenase system version 2.0 (United States Biochemical), which is a chain-termination method of DNA sequencing. The following procedure was used to read the nucleic acid sequence of the amplified products. 7 µl of PCR product was mixed with 2 µl shrimp alkaline phosphatase and 0.5 µl exonuclease I. The mixture was incubated at 37° C. for 15 min and then at 80° C. for 15 min. After addition of 1 µl of primer, the mixture was incubated at 100° C. for 3 min and then chilled on ice for 5 min. Next, the sample was incubated for 5 min at room temperature with the following additions: 2 µl 5× buffer, 1 µl DTT, 2 µl diluted dGTP, 0.5 µl ³⁵S-dATP, and 2 µl diluted Sequenase buffer. A 3.5 µl aliquot of the mixture was then added to 1 µl of one dideoxynTP. After 5 min at 37° C., 4 µl of stop solution was added to the tube. The products were run on a 6% polyacrylamide sequencing gel for 2–4 hr. Following this, the gel was dried on a BioRad gel dryer and exposed to film overnight. Film was developed on a Kodak M35A X-OMAT Processor. The method has been used successfully to duplicate the published sequence of the HoxA1 exons in samples from a number of controls. The film was developed the next afternoon, and the DNA sequence was read manually for comparison to the published Hox A1 sequence.

The nucleotide sequence from some patients, including the members of the family mentioned previously, showed the presence of two discrete bands at the same levels on the gel.

20

Example 3

Sequencing the PCR Products

Since sequencing PCR products allows the DNA sequence to be read from both alleles, a sequence with double bands suggests heterozygosity—that the two alleles are not the same and that two different sequences superimposed on one another are being read. Based on these results, the PCR products were cloned in order to get a cleaner sequence. Cloning separates the two alleles and allowed each to be individually sequenced to determine whether one or both alleles are abnormal.

The PCR products were cloned using Invitrogen's Zero Blunt PCR Cloning Kit. This kit is designed to clone blunt-ended PCR fragments, which can be generated by using a thermostable DNA polymerase with proofreading activity. Once the products were cloned, the clonal DNA was sequenced using the Sequenase version 2.0 chain-termination sequencing system. Each clone was sequenced in both 5' and 3' directions, and the reactions were run out for 6 hours on a 6% polyacrylamide sequencing gel.

Cloning allowed the determination that three out of four members of this family are indeed heterozygous for Hox A 1. The father and both children contain an identical mutation in the gene: a single base-pair change of A to G in the first exon of the gene; the mother's gene is normal. This mutation is dominant with variable penetrance. Sequences showing the mutation can be seen in FIG. 1. FIG. 1A shows the wild-type sequence. Substitution of guanine for adenine at this single location as shown in FIG. 1B causes an alteration in the resulting amino acid sequence, changing a histidine to an arginine.

Example 4

Restriction Analysis of PCR Products

The PCR products from this family were also subjected to restriction enzyme digestion to confirm the mutation. The enzyme Hph I recognizes the specific sequence 3' . . . CCACT(N₇) . . . 5'. When normal HoxA1 is digested with this enzyme, it will be cut; however, when mutated HoxA1 is digested, it will not be cut, because the recognition site has been changed by the mutation. This enzyme has been used to digest PCR products from this family and confirm that the mutation does indeed exist in the father and the children but not in the mother. This enzyme has been used to digest PCR products from approximately 100 controls, 36 parent pairs, 26 affected relatives, and 46 probands. In forty cases, the results of the restriction analysis has been compared to that from the sequencing reactions. The two methods gave identical results in every case.

Example 5

Sequencing of a Polymorphism in HoxB1

The sequence for the HoxB1 gene (accession number X16666) was obtained from the Entrez data base. From this sequence primers for the amplification of a 575 bp product of exon 1 by PCR were designed (Sense: 5'-GCATGGACTATAATAGGATG-3' (SEQ. ID. No. 9); Antisense: 5'-TCTTGGGTGGGTTTCTCTTA-3' (SEQ. ID. No. 10)). The final concentration of the following components were used in the amplification reaction: 1.5 U Taq polymerase; 200 µM each of dATP, dCTP, dGTP, dTTP; 1.5 mM MgCl₂; 0.4 mM of each sense and antisense primer; 50–100 ng DNA template; and distilled H₂O to a final volume of 25 µl. The Taq, dNTPs and MgCl₂ are supplied in a Ready-To-Go PCR Bead (Pharmacia 27-9555-01) and were used according to manufacturer's directions. The PCR reaction was carried out in a Perkin-Elmer 480 GeneAmp or a Perkin-Elmer 2400 thermocycler. Reaction conditions

were: denaturing for 1 minute at 94° C., and then 35 cycles of denaturing at 94° C. for 45 sec, annealing at 57° C. for 45 sec, and elongation at 72° C. for 45 sec. Resulting PCR product was analyzed on a 1% agarose gel and compare to a 100 bp ladder to determine the size of the product. Since the size of the product was as expected (575 bp) and somatic cell hybrid results indicated that the product is specific for chromosome 17 DNA samples from probands, family members and controls were amplified and sequenced using a radiolabeled terminator cycle sequencing kit (Amersham Life Science US79750). The sequencing reaction was ran on a 6% acrylamide sequencing gel (National Diagnostics) and exposed to Kodak Biomax MS X ray film for 24–48 hours. After developing the film, the resulting sequence was compared to the published sequence found in the Entrez data base.

Example 6

Association of the Newly-discovered Alleles with Autism Spectrum Disorders

Forty-six probands with autistic spectrum disorders and evidence of genetic causation were selected for analysis. Forty-three had one or more other affected family members and thirty-five had ear anomalies or neurological deficits consistent with malfunction of HoxA1 or its paralogs. For comparison, three other groups were tested:

1) An unstructured control group consisting of adults with no evidence of neurological abnormality collected from many different medical centers. These were mostly spouses of patients with late onset degenerative diseases of the nervous system. The purpose of this group was to determine the frequency of the alleles in the general population.

2) Parent controls—While each of the parents of a proband obviously transmits half of his or her genetic material to the proband, imaginary individuals with two alleles constructed from the untransmitted allele of each parent pair should give an accurate estimate of the frequency of the alleles in the study population, aside from those transmitted to the probands. Thus, the untransmitted alleles of the parent pair make a more stringent control, taking into account known and unknown structure in the local population.

3) Affected family members of probands—When they were available, the siblings, cousins, parents, or aunts and uncles of probands diagnosed with autism spectrum disorders or related symptoms (e.g. learning disabilities, language delays, neurological anomalies of the cranial nerves) were tested. If an allele is associated with autism, it should be more frequent in probands and affected family members than in historic or parent controls.

TABLE 1

Percent of individuals with polymorphic forms of HoxA1 and/or B1			
	HOXA1	HOXB1	HOXA1 or HOXB1
Historic controls (N = 101)	16	34	47
Parent controls (N = 36)	22	39	55†
Probands with ASD (N = 46)	35**	52*	80***
Other affected relatives (N = 24)	38*	42	75*

different from historical controls: * = $p < .05$, ** = $p < .01$, *** = $p < .001$

different from probands: † = $p < .05$

Table 1 demonstrates that parent controls are, indeed, similar to historic controls in their rates of the polymorphisms under study, while affected family members are similar to probands. This is especially true when the two functionally-related genes are combined. Eighty percent of

probands have one deviation from the normal sequence or the other, while only 47% of historical controls have an anomaly. Parent controls (untranslated alleles) match the historical controls in their rate of abnormal alleles, indicating that the local population is not structured differently from the general population in its rate of these alleles. In contrast, both probands ($\chi^2=14.83$, $p<0.001$) and other affected family members ($\chi^2=6.30$, $p<0.02$) differ significantly from historical controls. The probands differ significantly from the parent controls, as well $\chi^2=4.08$, $p<0.05$). The probands with genetic anomalies of HoxA1 or HoxB1 are concordant with the other affected members of the family in 18/22 cases ($\chi^2=17.82$, $p<0.001$). Finally, both the HoxA1 and HoxB1 polymorphisms are significantly associated with autism as judged by the Transmission Disequilibrium Test for Association (Spielman and Ewens, 1996), which compares the rate of transmission “into the disease” to the 50% rate one would expect in offspring of parents with the allele of interest. The χ^2 s for this test are: HoxA1=5.16, $p<0.05$; HoxB1=4.67, $p<0.05$.

In addition to the living probands, it was of interest to determine the genotype of the patient whose brain anatomy first suggested the involvement of the Hox genes in autism (Rodier et al., 1996). Genomic DNA was extracted from the autopsy tissue, and the patient was determined to have the B1 polymorphism (Stodgell et al., 1998).

One proband is homozygous for the less common allele of HoxA1, and he is severely affected. He was diagnosed early, at 21 months. None of the historic controls, and no parents, were homozygous for the polymorphism. Homozygosity of the HoxB1 polymorphism occurred in two historic controls, one affected parent, and in two severely-affected probands. Larger samples are needed to determine whether either polymorphism reduces viability. Three probands have both polymorphisms, and are severely disabled. The detection and description of the polymorphisms in the first exons of HoxA1 and HoxB1 and the progress of the association studies have been described in a book chapter and two abstracts (Rodier, 1998; Ingram et al., 1997; Stodgell et al., 1998).

Example 7

Identification of a Second Polymorphism in HoxA1

A third polymorphism has been detected in the homeobox region of HoxA1 in the second exon. The second exon cannot be amplified by PCR from the DNA of four probands indicating that an anomaly exists. This indicates that they are homozygous for a deviation from the published sequence on which the primers for the exon were based. PCR amplification yields suggest that about ten other probands are heterozygotes for this polymorphism of the second exon of HoxA1.

Additional primers have been developed that will allow complete sequencing of the altered region, which appears to be at the 3' end of the homeobox. Once the sequence is established, a test (such as the use of restriction length polymorphisms) can be developed to allow rapid evaluation of DNA samples. The degree of association of this polymorphism with autism spectrum disorders will then be studied in the same groups already evaluated for the others. Other studies in progress are designed to examine the second exon of HoxB1 and the non-coding regions of both genes.

Example 8

Identification of Additional Polymorphisms in HoxB1 and HoxD1 Associated with Autism

The procedures for evaluating the candidate gene HoxD1, as well as for finding additional polymorphisms in HoxA1

and HoxB1, will be the same as for those already identified in HoxA1 and HoxB1. Mutation detection in the coding sequence of these genes will consist of PCR amplification, cloning and sequencing. Mutation detection for the entire genes will include large deletion/insertion analysis by Southern blotting, analysis of 200-400 bp fragments by SSCP or heteroduplex analysis, and of course cloning and sequencing when heterozygosity becomes apparent for any region of the genes. *Current Protocols in Human Genetics* (John Wiley & Sons, Inc.), Chapter 7, "Searching Candidate Genes for Mutations."

Biological samples already isolated from patients with autism which did not show any abnormalities in HoxA1 or HoxB1 will be screened for polymorphisms in HoxD1.

Although preferred embodiments have been depicted and described in detail herein, it will be apparent to those skilled in the relevant art that various modifications, additions, substitutions, and the like can be made without departing from the spirit of the invention and these therefore are considered within the scope of the invention as defined in the claims which follow.

 SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 10

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1008 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

ATGGACAATG CAAGAATGAA CTCCTTCCTG GAATACCCCA TACTTAGCAG TGGCGACTCG      60
GGGACCTGCT CAGCCCGAGC CTACCCCTCG GACCATAGGA TTACAACTTT CCAGTCGTGC      120
GCGGTCAGCG CCAACAGTTG CGGCGGCGAC GACCGCTTCC TAGTGGGCAG GGGGGTGCAG      180
ATCGGTTGCG CCCACCACCA CCACCACCAC CACCATCACC ACCCCAGCC GGCTACCTAC      240
CAGACTTCGG GGAACCTGGG GGTGTCTTAC TCCCACTCAA GTTGTGGTCC AAGCTATGGC      300
TCACAGAACT TCAGTGCGCC TTACAGCCCC TACGCGTTAA ATCAGGAAGC AGACGTAAGT      360
GGTGGGTACC CCCAGTGC GC TCCCGCTGTT TACTCTGAA ATCTCTCATC TCCCATGGTC      420
CAGCATCACC ACCACCACCA GGGTTATGCT GGGGGCGCGG TGGGCTCGCC TCAATACATT      480
CACCCTCAT ATGGACAGGA GCACCAGAGC CTGGCCCTGG CTACGTATAA TAACCTCTTG      540
TCCCCTCTCC ACGCCAGCCA CCAAGAAGCC TGTCGCTCCC CCGCATCGGA GACATCTTCT      600
CCAGCGCAGA CTTTGTGACTG GATGAAAGTC AAAAGAAACC CTCCAAAAC AGGGAAGTT      660
GGAGAGTACG GCTACCTGGG TCAACCCAAC GCGGTGCGCA CCAACTTCAC TACCAAGCAG      720
CTCAGCGAAT TGGAGAAGGA GTTCCACTTC AACAAGTACC TGACGCGCGC CCGCAGGGTG      780
GAGATCGCTG CATCCCTGCA GCTCAACGAG ACCCAAGTGA AGATCTGGTT CCAGAACCGC      840
CGAATGAAGC AAAAGAAACG TGAGAAGGAG GGTCTCTTGC CCATCTCTCC GGCCACCCCG      900
CCAGGAAACG ACGAGAAGGC CGAGGAATCC TCAGAGAAGT CCAGCTCTTC GCCCTGCGTT      960
CCTTCCCGGG GGTCTCTTAC CTCAGACACT CTGACTACCT CCCACTGA      1008

```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 335 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

-continued

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

Met Asp Asn Ala Arg Met Asn Ser Phe Leu Glu Tyr Pro Ile Leu Ser
1      5      10      15
Ser Gly Asp Ser Gly Thr Cys Ser Ala Arg Ala Tyr Pro Ser Asp His
      20      25      30
Arg Ile Thr Thr Phe Gln Ser Cys Ala Val Ser Ala Asn Ser Cys Gly
      35      40      45
Gly Asp Asp Arg Phe Leu Val Gly Arg Gly Val Gln Ile Gly Ser Pro
      50      55      60
His His His His His His His His His His Pro Gln Pro Ala Thr Tyr
      65      70      75      80
Gln Thr Ser Gly Asn Leu Gly Val Ser Tyr Ser His Ser Ser Cys Gly
      85      90      95
Pro Ser Tyr Gly Ser Gln Asn Phe Ser Ala Pro Tyr Ser Pro Tyr Ala
      100     105     110
Leu Asn Gln Glu Ala Asp Val Ser Gly Gly Tyr Pro Gln Cys Ala Pro
      115     120     125
Ala Val Tyr Ser Gly Asn Leu Ser Ser Pro Met Val Gln His His His
      130     135     140
His His Gln Gly Tyr Ala Gly Gly Ala Val Gly Ser Pro Gln Tyr Ile
      145     150     155     160
His His Ser Tyr Gly Gln Glu His Gln Ser Leu Ala Leu Ala Thr Tyr
      165     170     175
Asn Asn Ser Leu Ser Pro Leu His Ala Ser His Gln Glu Ala Cys Arg
      180     185     190
Ser Pro Ala Ser Glu Thr Ser Ser Pro Ala Gln Thr Phe Asp Trp Met
      195     200     205
Lys Val Lys Arg Asn Pro Pro Lys Thr Gly Lys Val Gly Glu Tyr Gly
      210     215     220
Tyr Leu Gly Gln Pro Asn Ala Val Arg Thr Asn Phe Thr Thr Lys Gln
      225     230     235     240
Leu Thr Glu Leu Glu Lys Glu Phe His Phe Asn Lys Tyr Leu Thr Arg
      245     250     255
Ala Arg Arg Val Glu Ile Ala Ala Ser Leu Gln Leu Asn Glu Thr Gln
      260     265     270
Val Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Gln Lys Lys Arg Glu
      275     280     285
Lys Glu Gly Leu Leu Pro Ile Ser Pro Ala Thr Pro Pro Gly Asn Asp
      290     295     300
Glu Lys Ala Glu Glu Ser Ser Glu Lys Ser Ser Ser Pro Cys Val
      305     310     315     320
Pro Ser Pro Gly Ser Ser Thr Ser Asp Thr Leu Thr Thr Ser His
      325     330     335

```

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1008 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

-continued

ATGGACAATG CAAGAATGAA CTCCTTCCTG GAATACCCCA TACTTAGCAG TGGCGACTCG 60
 GGGACCTGCT CAGCCCCGAGC CTACCCCTCG GACCATAGGA TTACAACCTT CCAGTCGTGC 120
 GCGGTCAGCG CCAACAGTTG CGGCGCGGAC GACCGCTTCC TAGTGGGCAG GGGGGTGCAG 180
 ATCGGTTCGC CCCACCACCA CCACCACCAC CACCATCGCC ACCCCAGGCC GGCTACCTAC 240
 CAGACTTCCG GGAACCTGGG GGTGTCCTAC TCCCACTCAA GTTGTGGTCC AAGCTATGGC 300
 TCACAGAACT TCAGTGCGCC TTACAGCCCC TACGCGTTAA ATCAGGAAGC AGACGTAAGT 360
 GGTGGGTACC CCCAGTCGCG TCCCGCTGTT TACTCTGGAA ATCTCTCATC TCCCATGGTC 420
 CAGCATCACC ACCACCACCA GGGTTATGCT GGGGGCGCGG TGGGCTCGCC TCAATACATT 480
 CACCACTCAT ATGGACAGGA GCACCAGAGC CTGGCCCTGG CTACGTATAA TAACTCCTTG 540
 TCCCTCTCC ACGCCAGCCA CCAAGAAGCC TGTCGCTCCC CCGCATCGGA GACATCTTCT 600
 CCAGCGCAGA CTTTGTACTG GATGAAAGTC AAAAGAAACC CTCCCAAAC AGGAAAGTT 660
 GGAGAGTACG GCTACCTGGG TCAACCAAC GCGGTGCGCA CCAACTTCAC TACCAAGCAG 720
 CTCACGGAAC TGGAGAAGGA GTTCCACTTC AACAAGTACC TGACGCGCGC CCGCAGGGTG 780
 GAGATCGCTG CATCCCTGCA GCTCAACGAG ACCCAAGTGA AGATCTGGTT CCAGAACCGC 840
 CGAATGAAGC AAAAGAAACG TGAGAAGGAG GGTCTCTTGC CCATCTCTCC GGCCACCCCG 900
 CCAGGAAACG ACGAGAAGGC CGAGGAATCC TCAGAGAAGT CCAGCTCTTC GCCCTGCGTT 960
 CCTTCCCCGG GGTCTTCTAC CTCAGAACT CTGACTACCT CCCACTGA 1008

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 335 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Asp Asn Ala Arg Met Asn Ser Phe Leu Glu Tyr Pro Ile Leu Ser
 1 5 10 15
 Ser Gly Asp Ser Gly Thr Cys Ser Ala Arg Ala Tyr Pro Ser Asp His
 20 25 30
 Arg Ile Thr Thr Phe Gln Ser Cys Ala Val Ser Ala Asn Ser Cys Gly
 35 40 45
 Gly Asp Asp Arg Phe Leu Val Gly Arg Gly Val Gln Ile Gly Ser Pro
 50 55 60
 His His His His His His His Arg His Pro Gln Pro Ala Thr Tyr
 65 70 75 80
 Gln Thr Ser Gly Asn Leu Gly Val Ser Tyr Ser His Ser Ser Cys Gly
 85 90 95
 Pro Ser Tyr Gly Ser Gln Asn Phe Ser Ala Pro Tyr Ser Pro Tyr Ala
 100 105 110
 Leu Asn Gln Glu Ala Asp Val Ser Gly Gly Tyr Pro Gln Cys Ala Pro
 115 120 125
 Ala Val Tyr Ser Gly Asn Leu Ser Ser Pro Met Val Gln His His His
 130 135 140
 His His Gln Gly Tyr Ala Gly Gly Ala Val Gly Ser Pro Gln Tyr Ile
 145 150 155 160
 His His Ser Tyr Gly Gln Glu His Gln Ser Leu Ala Leu Ala Thr Tyr
 165 170 175

-continued

Asn Asn Ser Leu Ser Pro Leu His Ala Ser His Gln Glu Ala Cys Arg
 180 185 190
 Ser Pro Ala Ser Glu Thr Ser Ser Pro Ala Gln Thr Phe Asp Trp Met
 195 200 205
 Lys Val Lys Arg Asn Pro Pro Lys Thr Gly Lys Val Gly Glu Tyr Gly
 210 215 220
 Tyr Leu Gly Gln Pro Asn Ala Val Arg Thr Asn Phe Thr Thr Lys Gln
 225 230 235 240
 Leu Thr Glu Leu Glu Lys Glu Phe His Phe Asn Lys Tyr Leu Thr Arg
 245 250 255
 Ala Arg Arg Val Glu Ile Ala Ala Ser Leu Gln Leu Asn Glu Thr Gln
 260 265 270
 Val Lys Ile Trp Phe Gln Asn Arg Arg Met Lys Gln Lys Lys Arg Glu
 275 280 285
 Lys Glu Gly Leu Leu Pro Ile Ser Pro Ala Thr Pro Pro Gly Asn Asp
 290 295 300
 Glu Lys Ala Glu Glu Ser Ser Glu Lys Ser Ser Ser Ser Pro Cys Val
 305 310 315 320
 Pro Ser Pro Gly Ser Ser Thr Ser Asp Thr Leu Thr Thr Ser His
 325 330 335

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1021 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TGACGCATGG ACTATAATAG GATGAACTCC TTCTTAGAGT ACCCACTCTG TAACCGGGGA	60
CCCAGCGCCT ACAGCGCCCA CAGCGCCCA ACCTCCTTTC CCCCAGGCTC GGCTCAGGCG	120
GTTGACAGCT ATGCAAGCGA GGGCCGCTAC GGTGGGGGGC TGTCAGCCC TCGTTTCAG	180
CAGAACTCCG GCTATCCCGC CCAGCAGCCG CCTTCGACCC TGGGGGTGCC CTTCCCCAGC	240
TCCGCGCCCT CGGGGTATGC TCCTGCCGCC TGCAGCCCA GCTACGGGCC TTCTCAGTAC	300
TACCTCTGG GTCAATCAGA AGGAGACGGA GGCTATTTTC ATCCCTCAG CTACGGGGCC	360
CAGTAGGGG GCTTGTCCGA TGGCTACGGA GCAGGTGGAG CCGGTCCGGG GCCATATCCT	420
CCGCAGCATC CCCCTTATGG GAACGAGCAG ACCGCGAGCT TTGCACCGGC CTATGCTGAT	480
CTCTCTCCG AGGACAAGGA AACACCCTGC CCTTCAGAAC CTAACACCCC CACGGCCCGG	540
ACCTTCGACT GGATGAAGGT TAAGAGAAAC CCACCAAGA CAGCAAGGT GTCAGAGCCA	600
GGCTTGGGCT CGCCCAAGTG CCTCCGACC AACTTCACCA CAAGGCAGCT GACAGAACTG	660
GAAAAGGAGT TCCATTTCAA CAAGTACCTG AGCCGGGCCC GGAGGGTGGA GATTGCCGCC	720
ACCTTGAGC TCAATGAAC ACAGGTCAAG ATTTGGTTCC AGAACCGAGC AATGAAGCAG	780
AAGAAGCGCG AGCGAGAGGG AGGTCGGGTC CCCCAGCCC CACCAGGCTG CCCCAGGAG	840
GCAGCTGGAG ATGCTCAGA CAGTCGACA TGCACCTCCC CGGAAGCCTC ACCCAGCTCT	900
GTCACCTCCT GAATGAACC TAGCCACCAA TGGGGCTTCC AGGCACTGGA GCGCCCCAGT	960
CCAGCCCTAT CCCAGGCTCT CCAACCCAG GCCTGGCTTC ACTGCCTGGG ATCTCTAGGC	1020
T	1021

-continued

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 301 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```

Met Asp Tyr Asn Arg Met Asn Ser Phe Leu Glu Tyr Pro Leu Cys Asn
1           5           10          15
Arg Gly Pro Ser Ala Tyr Ser Ala His Ser Ala Pro Thr Ser Phe Pro
20          25          30
Pro Ser Ser Ala Gln Ala Val Asp Ser Tyr Ala Ser Glu Gly Arg Tyr
35          40          45
Gly Gly Gly Leu Ser Ser Pro Ala Phe Gln Gln Asn Ser Gly Tyr Pro
50          55          60
Ala Gln Gln Pro Pro Ser Thr Leu Gly Val Pro Phe Pro Ser Ser Ala
65          70          75          80
Pro Ser Gly Tyr Ala Pro Ala Ala Cys Ser Pro Ser Tyr Gly Pro Ser
85          90          95
Gln Tyr Tyr Pro Leu Gly Gln Ser Glu Gly Asp Gly Gly Tyr Phe His
100         105         110
Pro Ser Ser Tyr Gly Ala Gln Leu Gly Gly Leu Ser Asp Gly Tyr Gly
115         120         125
Ala Gly Gly Ala Gly Pro Gly Pro Tyr Pro Pro Gln His Pro Pro Tyr
130         135         140
Gly Asn Glu Gln Thr Ala Ser Phe Ala Pro Ala Tyr Ala Asp Leu Leu
145         150         155         160
Ser Glu Asp Lys Glu Thr Pro Cys Pro Ser Glu Pro Asn Thr Pro Thr
165         170         175
Ala Arg Thr Phe Asp Trp Met Lys Val Lys Arg Asn Pro Pro Lys Thr
180         185         190
Ala Lys Val Ser Glu Pro Gly Leu Gly Ser Pro Ser Gly Leu Arg Thr
195         200         205
Asn Phe Thr Thr Arg Gln Leu Thr Glu Leu Glu Lys Glu Phe His Phe
210         215         220
Asn Lys Tyr Leu Ser Arg Ala Arg Arg Val Glu Ile Ala Ala Thr Leu
225         230         235         240
Glu Leu Asn Glu Thr Gln Val Lys Ile Trp Phe Gln Asn Arg Arg Met
245         250         255
Lys Gln Lys Lys Arg Glu Arg Glu Gly Gly Arg Val Pro Pro Ala Pro
260         265         270
Pro Gly Cys Pro Lys Glu Ala Ala Gly Asp Ala Ser Asp Gln Ser Thr
275         280         285
Cys Thr Ser Pro Glu Ala Ser Pro Ser Ser Val Thr Ser
290         295         300

```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1030 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

-continued

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```

TGACGCATGG ACTATAATAG GATGAACTCC TTCTTAGAGT ACCCACTCTG TAACCGGGGA      60
CCCAGCGCCT ACAGCGCCCA CAGCGCCAC AGCGCCCAA CCTCCTTTCC CCCAAGCTCG      120
GCTCAGCGCG TTGACAGCTA TGCAAGCGAG GGCCGCTACG GTGGGGGGCT GTCCAGCCCT      180
GCGTTTCAGC AGAACTCCGG CTATCCCGCC CAGCAGCCGC CTTCGACCCT GGGGGTGCCC      240
TTCCCCAGCT CCGCGCCCTC GGGGTATGCT CCTGCCGCCT GCAGCCCCAG CTACGGGCCT      300
TCTCAGTACT ACCCTCTGGG TCAATCAGAA GGAGACGGAG GCTATTTTCA TCCCTCGAGC      360
TACGGGGCCC AGTAGGGGG CTTGTCCGAT GGCTACGGAG CAGGTGGAGC CGGTCCGGGG      420
CCATATCCTC CGCAGCATCC CCCTTATGGG AACGAGCAGA CCGCGAGCTT TGCACCGGCC      480
TATGCTGATC TCCTCTCCGA GGACAAGGAA ACACCCTGCC CTTCAGAACC TAACACCCCC      540
ACGGCCCGGA CCTTCGACTG GATGAAGGTT AAGAGAAACC CACCCAAGAC AGCGAAGGTG      600
TCAGAGCCAG GCCTGGGCTC GCCCAGTGGC CTCCGCACCA ACTTCACCAC AAGGCAGCTG      660
ACAGAAGTGG AAAAGGAGTT CCATTCAAC AAGTACCTGA GCCGGGCCCG GAGGGTGAG      720
ATTGCCGCCA CCCTGGAGCT CAATGAAACA CAGGTCAAGA TTTGGTTCCA GAACCGACGA      780
ATGAAGCAGA AGAAGCGCGA GCGAGAGGGA GGTCGGGTCC CCCCAGCCCC ACCAGGCTGC      840
CCCAAGGAGG CAGCTGGAGA TGCCTCAGAC CAGTCGACAT GCACCTCCCC GGAAGCCTCA      900
CCCAGTCTCT TCACCTCCTG AACTGAACCT AGCCACCAAT GGGGCTTCCA GGCCTGGAG      960
CGCCCCAGTC CAGCCCTATC CCAGGCTCTC CCAACCCAGG CCTGGCTTCA CTGCCTGGGA     1020
TCTCTAGGCT                                     1030

```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 304 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```

Met Asp Tyr Asn Arg Met Asn Ser Phe Leu Glu Tyr Pro Leu Cys Asn
1           5           10          15
Arg Gly Pro Ser Ala Tyr Ser Ala His Ser Ala His Ser Ala Pro Thr
20          25          30
Ser Phe Pro Pro Ser Ser Ala Gln Ala Val Asp Ser Tyr Ala Ser Glu
35          40          45
Gly Arg Tyr Gly Gly Gly Leu Ser Ser Pro Ala Phe Gln Gln Asn Ser
50          55          60
Gly Tyr Pro Ala Gln Gln Pro Pro Ser Thr Leu Gly Val Pro Phe Pro
65          70          75          80
Ser Ser Ala Pro Ser Gly Tyr Ala Pro Ala Ala Cys Ser Pro Ser Tyr
85          90          95
Gly Pro Ser Gln Tyr Tyr Pro Leu Gly Gln Ser Glu Gly Asp Gly Gly
100         105         110
Tyr Phe His Pro Ser Ser Tyr Gly Ala Gln Leu Gly Gly Leu Ser Asp
115         120         125
Gly Tyr Gly Ala Gly Gly Ala Gly Pro Gly Pro Tyr Pro Pro Gln His

```

-continued

130	135	140
Pro Pro Tyr Gly Asn Glu Gln Thr Ala Ser Phe Ala Pro Ala Tyr Ala		
145	150	155
Asp Leu Leu Ser Glu Asp Lys Glu Thr Pro Cys Pro Ser Glu Pro Asn		
	165	170
Thr Pro Thr Ala Arg Thr Phe Asp Trp Met Lys Val Lys Arg Asn Pro		
	180	185
Pro Lys Thr Ala Lys Val Ser Glu Pro Gly Leu Gly Ser Pro Ser Gly		
	195	200
Leu Arg Thr Asn Phe Thr Thr Arg Gln Leu Thr Glu Leu Glu Lys Glu		
	210	215
Phe His Phe Asn Lys Tyr Leu Ser Arg Ala Arg Arg Val Glu Ile Ala		
	225	230
Ala Thr Leu Glu Leu Asn Glu Thr Gln Val Lys Ile Trp Phe Gln Asn		
	245	250
Arg Arg Met Lys Gln Lys Lys Arg Glu Arg Glu Gly Gly Arg Val Pro		
	260	265
Pro Ala Pro Pro Gly Cys Pro Lys Glu Ala Ala Gly Asp Ala Ser Asp		
	275	280
Gln Ser Thr Cys Thr Ser Pro Glu Ala Ser Pro Ser Ser Val Thr Ser		
	290	295

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = "primer"

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GCATGGACTA TAATAGGATG

20

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: other nucleic acid
 (A) DESCRIPTION: /desc = "primer"

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TCTTGGGTGG GTTCTCTTA

20

55

What is claimed:

1. A method for screening subjects for genetic markers associated with autism, comprising:

isolating a biological sample from a mammal; and
 testing the sample or genetic material isolated from the sample for a polymorphism in a Hox A1 or B1 coding sequence which is a genetic marker for autism.

2. The method according to claim 1, wherein the biological sample is selected from the group consisting of blood, saliva, amniotic fluid, and tissue.

3. The method according to claim 2, wherein the biological sample is blood.

4. The method according to claim 1, wherein the mammal is a human.

5. The method according to claim 4, wherein the biological sample is isolated from developmentally disabled children.

6. The method according to claim 4, wherein the biological sample is isolated from parents or relatives of developmentally disabled children.

7. The method according to claim 4, wherein the biological sample is isolated from children and said method further comprises:

37

early behavior training for children having genetic markers associated with autism.

8. The method according to claim 1, wherein the polymorphism is located in the homeobox.

9. The method according to claim 1, wherein the coding sequence has a single base substitution resulting in an amino acid substitution.

10. The method according to claim 9, wherein the amino acid substitution is an arginine for a histidine.

11. The method according to claim 10, wherein the coding sequence has an insertion.

12. The method according to claim 11, wherein the insertion is 5' ACAGCGCCC-3'.

13. The method according to claim 1, wherein the coding sequence has a polymorphism selected from the group consisting of a single base substitution resulting in an amino acid substitution, a single base substitution resulting in a translational stop, an insertion, a deletion, and a rearrangement.

14. The method according to claim 1, wherein the polymorphism alters the sequence of the polypeptide encoded by the coding sequence.

15. The method according to claim 1, wherein said screening for mutated nucleic acids is carried out by a method selected from the group consisting of direct sequencing of nucleic acids, single strand polymorphism assay, restriction fragment length polymorphism assay, ligase chain reaction, enzymatic cleavage and southern hybridization.

16. The method according to claim 15, wherein said screening is carried out by direct sequencing of nucleic acids.

17. The method according to claim 15, wherein said screening is carried out by single strand polymorphism assay.

18. The method according to claim 15, wherein said screening is carried out by restriction fragment length polymorphism assay.

38

19. The method according to claim 15, wherein said screening is carried out by ligase chain reaction.

20. The method according to claim 15, wherein said screening is carried out by enzymatic cleavage.

21. The method according to claim 15, wherein said screening is carried out by southern hybridization.

22. The method according to claim 15, wherein the nucleic acid is a deoxyribonucleic acid.

23. The method according to claim 15, wherein the nucleic acid is a messenger ribonucleic acid.

24. An isolated nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO: 1, wherein the nucleic acid molecule comprises a single base substitution at nucleotide 218.

25. The isolated nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:5, wherein the nucleic acid molecule comprises an insertion between nucleotides 88 and 89.

26. The isolated nucleic acid molecule according to claim 25, wherein the insertion is 5'-ACAGCGCCC-3'.

27. An isolated nucleic acid molecule consisting of at least 15 contiguous nucleotides of the coding sequence set forth in SEQ ID NO:5 wherein the molecule comprises an insertion between nucleotides 88 and 89 in SEQ ID NO:5 and wherein the molecule specifically binds to a HoxA1 or HoxB1 coding sequence but does not bind to other coding sequences.

28. An isolated nucleic acid molecule consisting of at least 15 contiguous nucleotides of the coding sequence set forth in SEQ ID NO: 1 wherein the molecule comprises a single base substitution at nucleotide 218 and wherein the molecule specifically binds to a HoxA1 or HoxB1 coding sequence but does not bind to other coding sequences.

29. The method according to claim 1 wherein the coding sequence has a mutation in an exon.

* * * * *



US006162604A

United States Patent [19]**Jacob**[11] **Patent Number:** **6,162,604**[45] **Date of Patent:** **Dec. 19, 2000****[54] METHODS FOR DETERMINING GENETIC PREDISPOSITION TO AUTOIMMUNE DISEASES BY GENOTYPING APOPTOTIC GENES****[76] Inventor:** **Chaim O. Jacob**, 2110 Beverwil Dr., Los Angeles, Calif. 90034[21] Appl. No.: **09/283,040**[22] Filed: **Apr. 1, 1999**[51] **Int. Cl.**⁷ **C12Q 1/68; C12P 19/34; C07H 21/04**[52] **U.S. Cl.** **435/6; 435/91.2; 536/24.31; 536/24.33**[58] **Field of Search** **435/6, 810; 536/24.31, 536/24.33****[56] References Cited
PUBLICATIONS**

Tsujimoto et al., "DNA rearrangements in human follicular lymphoma can involve the 5' or 3' region of the bcl-2 gene", Proceedings of the National Academy of Sciences, USA, vol. 84, pp. 1329-1331, Mar. 1987.

Platzer et al., "Up-regulation of monocytic IL-10 by tumor necrosis factor-alpha and cAMP elevating drugs", International Immunology, vol. 7 (4), pp. 517-523, Dec. 1994.

Alderson et al., "Fas ligand Mediates Activation-induced cell death in human T lymphocytes", Journal of Experimental Medicine, vol. 181, pp. 71-77, Jan. 1995.

Harley et al., "Lupus and Interleukin 10", Journal of Rheumatology, vol. 24 (12), pp. 2273-2275, 1997.

Aringer et al., "High levels of bcl-2 protein in circulating T lymphocytes, but not B lymphocytes of patients with systemic lupus erythematosus", Arthritis and Rheumatism, vol. 37 (10), pp. 1423-1430, Oct. 1994.

Donner et al., "CTLA4 alanine-17 confers genetic susceptibility to Graves' disease and to Type 1 Diabetes Mellitus", Journal of clinical endocrinology and metabolism, vol. 82 (1), pp. 143-146, 1997.

Huang et al., "Identification and characterization of polymorphism in the promoter region of the human Apo-1/Fas (CD95) gene", Molecular Immunology, vol. 34 (8-9), pp. 577-582, 1997.

Eskdale, et al., Association between polymorphisms at the human IL-10 locus and systemic lupus erythematosus, *Tissue Antigens* 49:635-639 (1997).

Polymeropoulos, et al., Tetranucleotide repeat polymorphism at the human c-fes/fps proto-oncogene (FES), *Nucleic Acids Research* 19(14):4018 (1991).

Tan et al., The 1982 Revised Criteria For The Classification of Systemic Lupus Erythematosus, *Arthritis and Rheumatism* 25(11):1271-1277 (1982).

Turner, et al., An Investigation of Polymorphism in the Interleukin-10 Gene Promoter, *European Journal of Immunogenetics* 24:1-8 (1997).

Primary Examiner—Jeffrey Fredman**Assistant Examiner**—Arun Chakrabarti**Attorney, Agent, or Firm**—Knobbe Martens Olson & Bear, LLP**[57] ABSTRACT**

Genetic markers associated with programmed cell death were characterized and their extent of polymorphism in normal populations was determined allowing for a method for determining genetic predisposition to SLE and other autoimmune diseases by genotyping. The allelic distribution of these gene markers in a large Mexican American SLE cohort and ethnically matched controls was determined. The results were that bcl-2, Fas-L, and IL-10 loci showed significantly different allelic distribution in SLE patients compared with controls, indicating an association between these genes and SLE. The method allows for determining the presence of these alleles. Alone, the presence of each of these alleles is associated with a moderate increase in SLE risk, while the occurrence of these alleles together increases the odds of developing SLE by more than 40-fold.

10 Claims, No Drawings

METHODS FOR DETERMINING GENETIC PREDISPOSITION TO AUTOIMMUNE DISEASES BY GENOTYPING APOPTOTIC GENES

FIELD OF THE INVENTION

This invention relates generally to methods for determining predisposition to systemic lupus erythematosus (SLE) and other autoimmune diseases by genotyping IL-10, bcl-2, FAS ligand (FAS-L) and other apoptotic genes. More specifically, the bcl-2, Fas-L, and IL-10 loci showed significantly different allelic distribution in SLE patients compared with controls, indicating an association between these genes and SLE. Additionally, further analysis revealed a synergistic effect between susceptibility alleles of the bcl-2 and IL-10 genes in determining disease susceptibility.

BACKGROUND OF THE INVENTION

Systemic lupus erythematosus (SLE) is considered to be the prototype of human autoimmune diseases. It is a disorder of generalized autoimmunity characterized by multisystem organ involvement, polyclonal B cell activation, and the production of autoantibodies against nuclear, cytoplasmic, and cell surface antigens. Autoreactive B and T lymphocytes can be found in healthy individuals as well, but their numbers are tightly regulated by a process of programmed cell death (apoptosis), which is crucial in the establishment of self-tolerance. Tolerance to self antigens can fail and can result in autoimmunity if there is a defect in the process of elimination of these cells.

SLE, as well as most other autoimmune diseases is difficult to diagnose. A strict definition of SLE patients included 4 or more of the 11 ACR revised criteria for SLE, eliminating LE cells but adding anticardiolipin antibodies and lupus anticoagulant as criteria (Tan, et al. *Arthritis Rheum.* 1982;25:1271-7). Often, a patient that is going to develop SLE will be kept off of treatment because they only show two or three of the criteria. One of the major tests, the ANA test (anti-nuclear antibodies), tests for the presence of these antibodies. However, 15-20% of those individuals with a positive ANA will never develop disease. The inadequacy of definitive tests for the diagnosis of autoimmune diseases is a recurrent theme. For this reason, treatment is often not started until disease is too far along and irreversible damage has occurred. Therefore development of a test for the diagnosis and susceptibility to autoimmune diseases could have a profound effect on the outcome of the disease and the patient's quality of life.

Several lines of evidence suggest that dysfunctional programmed cell death (apoptosis) might be involved in the pathogenesis of SLE and other autoimmune diseases. It has been postulated that in SLE, dysfunction of apoptosis could result in the inappropriate longevity of autoreactive B lymphocytes, allowing autoantibody levels to reach pathogenic thresholds and breakdown of self tolerance. Defective apoptosis of autoreactive lymphocytes is an attractive mechanism contributing to SLE, primarily because defects in either the apoptosis-promoting Fas gene or its ligand Fas-L (CD95L) accelerates autoimmunity in mouse strains (MRL-lpr/lpr and C3H-gld/gld, respectively) that exhibit SLE-like diseases. Furthermore, studies reveal links between autoimmunity and several other gene products involved in apoptosis. The bcl-2 gene enhances lymphocyte survival by inhibiting or delaying apoptosis. Transgenic mice overexpressing bcl-2 in their B cells show polyclonal B cell expansion and extended survival in vitro. After a few

months, these mice developed an autoimmune syndrome resembling SLE.

Interleukin-10 (IL-10) is a pleiotropic cytokine that regulates many immune and inflammatory responses. Among other activities, this cytokine increases the survival of activated lymphocytes. Furthermore, administration of recombinant IL-10 to lupusprone (New Zealand black×New Zealand white)_{F₁} [(NZB×NZW)_{F₁}] mice accelerates the development of autoimmunity. CTLA-4 is an additional gene involved in apoptosis that has been suggested to be associated with autoimmune disease development. CTLA4 can mediate antigen-specific apoptosis and appears to be part of a distinct signaling pathway capable of clonally deleting previously activated human T lymphocytes. CTLA-4 also warrants further study because it may be a candidate gene in more than one autoimmune disease. CTLA-4 was reported to be associated with 2 autoimmune diseases, Grave's disease and insulin-dependent diabetes mellitus.

In a recent publication, Eskdale et al (*Tissue Antigens* 1997;49:635-9) have shown an association between an IL-10 microsatellite polymorphism and SLE in a Caucasian population. In this study a group of 56 Caucasian SLE patients from Great Britain were compared with 102 ethnically matched controls. However, because of the moderate sample size, the results were considered only as a framework for further study.

SUMMARY OF THE INVENTION

One object of the present invention is a method for determining predisposition to an autoimmune disease by obtaining a patient sample, amplifying at least two apoptotic loci, and determining whether the disease-specific allele is present.

A further embodiment includes identifying the disease-specific alleles by comparing the most abundant allele found in patients with disease to normal individuals. Preferably the apoptotic gene loci are selected from the group consisting of IL-10, bcl-2, Fas-L, and CTLA-4. Preferably, the IL-10 disease-associated allele is PCR amplified with the primers comprising SEQ ID NO:1 and SEQ ID NO:2, the bcl-2 disease-associated allele is PCR amplified with the primers comprising SEQ ID NO:3 and SEQ ID NO:4, the Fas-L disease-associated allele is PCR amplified with the primers comprising SEQ ID NO:5 and SEQ ID NO:6, and the CTLA-4 disease-associated allele is PCR amplified with the primers comprising SEQ ID NO:7 and SEQ ID NO:8. In a further preferred embodiment, the disease-associated allele is identified by size or sequence. Preferably, the disease is selected from the group consisting of; systemic lupus erythematosus, thyroid autoimmunity syndromes, insulin dependent diabetes mellitus, inflammatory bowel disease, rheumatoid arthritis and other arthritides.

A further object of the invention is a kit for determining predisposition to an autoimmune disease comprising the method of claim 1.

A further object of the invention is a method for producing a diagnostic test for predisposition to an autoimmune disease which involves obtaining a patient sample, PCR amplifying at least two apoptotic loci, and identifying the disease-specific alleles by comparison to normal individuals, finally determining whether the disease-specific allele is present in a test patient's sample.

DETAILED DESCRIPTION OF THE INVENTION

Because bcl-2, Fas-L, CTLA-4, and IL-10 participate in apoptosis, and because of the evidence suggesting that these

genes may be involved in the pathogenesis of SLE, we tested whether there is an association between these genes and SLE in humans.

The method of the present invention comprises a technique for determining the presence of disease-associated alleles of apoptotic genes and analyzing whether they show predisposition to autoimmune diseases.

Further features and advantages will become apparent to those of skill in the art in view of the Detailed Description of the Invention which follows, when considered together with the attached claims.

Although other materials and methods can be used in the practice or testing of the present invention, a method is now described. Examples 1-3 show how a method for determining predisposition to an autoimmune disease can be developed.

EXAMPLE 1

Characteristics of the Study Population

Patients in this study were from the University of Southern California (USC) School of Medicine clinics who were confirmed to have met the American College of Rheumatology (ACR) criteria for SLE. A strict definition of SLE patients included 4 or more of the 11 ACR revised criteria for SLE, eliminating LE cells but adding anticardiolipin antibodies and lupus anticoagulant as criteria (Tan, et al. *Arthritis Rheum.* 1982:25:1271-7).

We used semistructured personal or telephone interviews to obtain a complete family history of each SLE patient and control subject. Through these interviews, data were collected describing a fixed family structure (proband's grandparents, parents, siblings, and offspring, as well as siblings and offspring of both of the proband's parents). Information regarding the birthplace of the probands, their parents, and their grandparents was also obtained. Whenever possible, we obtained family history information about the probands from an additional source (usually, a parent of the subject).

For the purpose of this study, Mexican Americans were defined as individuals born in Mexico or the US whose grandparents from both the mother's and the father's side were born in Mexico. Controls were defined as Mexican American subjects who did not have SLE or any other autoimmune disease and whose family lacked any autoimmune disease history. The study protocol was approved by the Institutional Review Board of the USC School of Medicine.

EXAMPLE 2

Genotypic Analysis of IL-10, bcl-2, Fas-L, and CTLA-4

Blood samples were collected from all participants, and genomic DNA was extracted from the peripheral blood mononuclear cells by standard procedures. To obtain genotypes of the IL-10, bcl-2, and Fas-L, short tandem repeat sequences (microsatellites) within the noncoding regions of these genes were identified and used as intragenic markers. The Fas-L (TG)_n tandem repeat was identified in the 3'-untranslated region of the gene, ~600 basepairs after the stop codon, while the IL-10 (CA)_n microsatellite is located ~1 kb 5' to the ATG codon. The CTLA-4 dinucleotide repeat begins at bp 642 of exon 3 of the human CTLA-4 gene (Polymopoulos, et al. *Nucleic Acids Res* 1991:19:4018), and the bcl-2 (AC)_n microsatellite is located 570 bp 5' to the

ATG codon. (The IL-10, bcl-2, and Fas-L gene sequences can be found using Genomic Data Base accession numbers X78437, X51898, and GenBank number U08137, respectively).

To amplify or image these loci, PCR was performed as follows: unique oligonucleotide sequences flanking each microsatellite were designed as primers, one of which was labeled with a fluorescent dye and used in the polymerase chain reaction (PCR). The oligonucleotides flanking the IL-10 (CA)_n repeat were the 5' primer 5'-GCA ACA CTC CTC GTC GCA AC-3' (SEQ ID NO:1) and the 3' primer, tagged with the fluorescent dye 6FAM, 5'-CCT CCC AAA GAA GCC TTA GTA G-3' (SEQ ID NO:2). The oligonucleotides flanking the bcl-2 (AC)_n repeat were the 5' primer, tagged with the fluorescent dye TET, 5'-CGT GTA CAC ACT CTC ATA CAC GGC T-3' (SEQ ID NO:3) and the 3' primer 5'-GGG AGG GTG CGC CAT GAA AA-3' (SEQ ID NO:4). The oligonucleotides flanking the Fas-L (TG)_n repeat were the 5' primer, tagged with the fluorescent dye 6FAM, 5'-CA CTT CT AAA TGC ATA TCC TGA GCC-3' (SEQ ID NO:5) and the 3' primer 5'-TGT CAG GAA GCA TTC AAA ATC TTG ACC A-3' (SEQ ID NO:6).

For CTLA-4, we used an (AT)_n microsatellite marker previously described (Polymopoulos, et al. *Nucleic Acids Res* 1991:19:4018). The oligonucleotides flanking the CTLA-4 (AT)_n repeat were the 5' primer, tagged with the fluorescent dye TET, 5'-GCC AGT GAT GCT AAA GGT TG-3' (SEQ ID NO:7) and the 3' primer 5'-AAC ATA CGT GGC TCT ATG CA-3' (SEQ ID NO:8).

PCR amplification was carried out using 40 ng of genomic DNA. The reaction conditions consisted of 0.5 μ M of each primer (labeled and unlabeled), 10 mM Tris HCl, pH 8.8, 50 mM KCl, 1.5 mM MgCl₂, 50 μ M of each dNTP, and 0.2 units of Taq polymerase. For IL-10, the samples were processed through 30 cycles of 30 seconds at 94° C., 30 seconds at 57° C., and 30 seconds at 72° C. For CTLA-4 the conditions were 30 seconds at 94° C., 120 seconds at 55° C., and 30 seconds at 72° C.

A "touchdown" PCR assay for Fas-L and bcl-2 polymorphism was performed to circumvent spurious priming during amplification. The initial annealing temperature was 66° C.; subsequent annealing temperatures were decreased by 1° C. every cycle to a "touchdown" annealing temperature of 55° C., at which 30 cycles of 1 minute at 94° C., 1 minute at 55° C., and 1 minute at 72° C. were performed.

Aliquots of the PCR product were electrophoresed on a 377 Prism ABI sequencer (Applied Biosystems, Foster City, Calif.), and the fluorescent signal was recorded and analyzed by the Genescan software (Applied Biosystems). Different fluorescent dyes were plotted separately, and the sizes of the fluorescent peaks were estimated in basepairs by reference to the in-lane size standard Tamra 500 (Applied Biosystems). Microsatellite alleles were classified automatically according to their size using the Genotyper software (Applied Biosystems). For quality control to ensure reproducibility of allele assignments between gels, 1 lane in each gel was loaded with a sample that had previously been genotyped. Each lane of the sequencing gel was loaded with the internal size marker labeled with Tamra 500. In addition to the automated allele calling, we performed manual surveillance of every genotype.

Although Example 2 makes use of PCR amplification to determine sequence length polymorphisms, one of skill in the art can readily identify other methods for the purpose of identifying disease-specific alleles. Single point mutations can also be readily identified using a number of techniques

well known to those having ordinary skill in the art. Examples of such methods to identify small allelic differences include FISH (Fluorescence In Situ Hybridization), RFLP (Restriction fragment length polymorphism), TGGE (temperature gradient gel electrophoresis) and SSCP (single-strand conformation polymorphism), each of which can be used to identify differences in DNA or RNA. Pure hybridization methods, such as Southern blotting or DNA chip technology, can also be used. Alternatively differences in the protein product could be imaged or identified using such techniques as Western blotting, ELISA, or even enzymatic assays.

EXAMPLE 3

Statistical Analysis

Associations between loci and the presence of SLE were tested by fitting a logistic regression model to the data. Genotypes at each locus were coded assuming a multiplicative model for allelic effects. Under this model, the odds ratio for a person with alleles a_i and a_j is given by $e^{b_i} e^{b_j}$, where b_i and b_j are regression coefficients corresponding to a_i and a_j , respectively. For each locus, alleles that occurred in <3 subjects were eliminated from the analysis. The likelihood ratio test was used as a global test of association between each locus and the presence of SLE.

Pairwise interactions between IL-10, Fas-L, and bcl-2 alleles were modeled using a departure from a multiplicative model for the corresponding joint locus effects. At each locus, a genotype for each subject was coded based on the presence or absence of at least 1 copy of the corresponding high-risk allele. Using a logistic regression model, the likelihood ratio test was used to determine whether each interaction significantly improved the model fit compared with a model including only the main effects on the 2 component loci.

A significance level of 0.05 was used in all global testing. A bonferroni adjustment was used in determining the significance of individual alleles. P values are reported for all tests so that the reader may independently assess statistical significance.

EXAMPLE 4

Association of SLE with Apoptotic Markers

Highly polymorphic short tandem repeat sequences (microsatellites) within the noncoding regions of the Fas-L, bcl-2, and IL-10 genes were identified and characterized as part of the present study, and were used as markers (see EXAMPLE 1). The polymorphism information content scores were 0.72 for IL-10, 0.47 for bcl-2, 0.59 for Fas-L, and 0.83 for CTLA-4.

The allelic distribution of these microsatellites was determined in several distinct ethnic populations, including Caucasian Americans, African Americans, Chinese Americans, and Mexican Americans, and showed a significant variation among these ethnic groups. For example, Table 1 illustrates significant variation in bcl-2 allele frequencies among normal individuals belonging to 4 major ethnicities in the US. The global likelihood ratio, testing for differences in allelic distribution at bcl-2 among the 4 American populations, was $\chi^2=149.7$ (degrees of freedom [df]=18, $P=0.001$). Similar ethnic variation in marker allele frequencies was found in the other genes tested in this study. The allele frequencies observed in control populations conform to Hardy-Weinberg expectations.

TABLE 1

Allele distribution of bcl-2 microsatellite in various American populations*				
Allele (bp)	Frequency			
	CA (2n = 160)	AA (2n = 172)	MA (2n = 440)	ChA (2n = 100)
187	—	—	0.002	—
191	0.063	0.169	0.177	0.260
193	0.025	0.081	0.048	0.030
195	0.831	0.430	0.700	0.470
197	0.025	0.047	0.029	0.120
199	0.006	0.041	0.009	—
201	0.031	0.180	0.029	0.120
203	0.019	0.047	—	—
207	—	0.005	—	—

*2n = number of chromosomes scored to determine allele frequencies; CA = Caucasian Americans; AA = African Americans; MA = Mexican Americans; ChA = Chinese Americans; bp = basepairs.

Since SLE itself occurs at a higher frequency in certain ethnic populations than in others, an association between the disease and a gene marker might occur as a statistical artifact in the mixed population. To minimize this potential problem of population stratification, we decided to focus the study on one ethnic population in detail. We focused on Mexican Americans since they comprise the majority of SLE patients in our center. The data presented below were obtained from 158 Mexican American SLE patients and 223 ethnically matched control subjects. Selected clinical characteristics of the SLE patients in the study are shown in Table 2. Both cohorts (SLE patients and control subjects) were not significantly different in age and sex distribution.

TABLE 2

Selected clinical characteristics of the study population*		
Characteristic	SLE patients (n = 158)	Control subjects (n = 223)
Age, mean \pm SD years	34.2 \pm 11.6	35.4 \pm 12.7
Female, %	90.5	86
ANA positive, %	100	—
Anti-dsDNA positive, %	61	—
Renal involvement, %	35	—
CNS involvement, %	9	—

*SLE = systemic lupus erythematosus; ANA = antinuclear antibodies; anti-dsDNA = anti-double-stranded DNA antibodies; CNS = central nervous system.

The allelic distributions of microsatellite markers of the bcl-2, IL-10, and Fas-L genes in SLE cases and in ethnically matched controls are summarized in Table 3. Associations between these loci and the presence of SLE were tested by fitting a logistic regression model to the data (see EXAMPLE 1).

Bcl-2 We identified 9 distinct alleles of the bcl-2 gene; the most frequent allele in the controls was 195-bp long (Bcl-2₁₉₅). The global likelihood ratio statistic, which tests for a difference in allelic distribution at bcl-2 between cases and controls, was $\chi^2=34.95$ (df=5, $P=0.0001$), indicating a definite association between the bcl-2 gene and SLE.

IL-10 Regarding the IL-10 gene, 10 distinct alleles were found in Mexican Americans. The most common allele in the control population was 125-bp long (IL-10₁₂₅). The test of association of this gene with SLE gave $\chi^2=33.20$ (df=8, $P=0.0001$), indicating an association.

Fas-L The Fas-L intragenic marker showed 7 distinct alleles; allele 241 was the most common in the control

population. The global likelihood ratio test statistic for Fas-L was $\chi^2=23.99$ (df=6, $P=0.0005$), suggesting an association between Fas-L and SLE as well.

TABLE 3

Allele distribution of the intragenic markers of IL-10, bcl-2, and Fas-L in Mexican American SLE patients and normal controls*							
IL-10 allele frequency			bcl-2 allele frequency			Fas-L allele frequency	
Allele (bp)	Cases (2n = 316)	Controls (2n = 440)	Allele (bp)	Cases (2n = 312)	Controls (2n = 440)	Allele (bp)	Cases (2n = 298)
121	0.003	0.009	187	—	0.002	233	0.013
123	0.025	0.050	189	0.006	—	235	0.013
125	0.363	0.493	191	0.187	0.177	237	0.024
127	0.199	0.081	193	0.135	0.048	239	0.289
129	0.107	0.066	195	0.548	0.700	241	0.527
131	0.067	0.064	197	0.042	0.029	243	0.128
133	0.136	0.127	199	0.013	0.009	245	0.007
135	0.073	0.098	201	0.048	0.029		
137	0.022	0.006	203	0.026	—		
139	—	0.004					

*2n = number of chromosomes scored to determine allele frequencies.
SLE = systemic lupus erythematosus.

CTLA-4 The CTLA-4 marker, however, showed no association with SLE. As shown in Table 4, the CTLA-4 marker had 19 distinct alleles in the Mexican American population. The likelihood ratio test result between cases and controls was $\chi^2=19.5$ (df=13, $P=0.1074$). (Five alleles of the CTLA-4 occurred so rarely in the data set that accurate estimates of their odds ratios could not be calculated. These alleles were left out of the analysis.)

To further investigate the significant associations, we performed additional analyses to determine which allele(s) of bcl-2, Fas-L, and IL-10 were associated with SLE. Table 5 summarizes the odds ratio (OR) and 95% confidence intervals (95% CI) for the effect of each allele relative to a baseline allele. The Bcl-2₁₉₃ and Bcl-2₂₀₁ alleles were associated with increased odds of developing SLE (OR 5.61, $P=0.0001$ and OR 3.15, $P=0.006$ per allele copy, respectively, compared with Bcl-2₁₉₅).

With regard to IL-10, only the IL-10₁₂₇ was associated with increased odds of developing SLE (OR 2.81 per allele copy, as compared with IL-10₁₂₅, $P=0.0001$). The Fas-L₂₃₉ allele was associated with increased odds of developing SLE (OR 1.69 per allele copy), as compared with the Fas-L₂₄₁ allele ($P=0.001$). As expected, the CTLA-4 gene showed no specific allele association with SLE (Table 5).

EXAMPLE 5

Synergistic Association of IL-10 and Bcl-2 Alleles

We next explored the possibility that synergistic effects between these loci may increase the risk of developing SLE. To this end, a departure from a multiplicative model for corresponding allelic effects was tested. To minimize the number of tests, we focused on single high-risk allele at each locus: allele 193 at bcl-2, allele 127 at IL-10, and allele 239 at Fas-L (see Table 5). The interaction tests are summarized in Table 6. We found no significant interaction between IL-10 and Fas-L, or between Fas-L and bcl-2. However, surprisingly there was significant interaction between the IL-10₁₂₇ allele and the Bcl-2₁₉₃ allele ($P=0.004$). Of 23 subjects that carried both the Bcl-2₁₉₃ and IL-10₁₂₇ alleles, 22 had SLE. While a person carrying either the IL-10₁₂₇ or the Bcl-2₁₉₃ allele only had an OR of ~2, a person carrying

both the IL-10₁₂₇ and the Bcl-2₁₉₃ susceptibility alleles together had an OR of 40.71 (Table 7).

TABLE 4

Allele distribution of CTLA-4 in Mexican American SLE patients and normal controls.			
Allele (bp)	Cases (2n = 250)	Controls (2n = 446)	
88	0.564	0.581	
94	0.004	0.006	
96	—	0.006	
102	0.012	0.004	
104	0.076	0.058	
106	0.188	0.222	
108	0.036	0.027	
110	0.036	0.009	
112	0.008	0.004	
114	0.008	0.006	
116	—	0.004	
118	0.012	0.070	
120	0.016	0.002	
122	0.008	0.004	
124	0.016	0.012	
126	0.008	0.011	
128	0.004	0.006	
130	0.004	0.020	
132	—	0.006	

*2n = number of chromosomes scored to determine allele frequencies.
SLE = systemic lupus erythematosus.

TABLE 5

Association between SLE and specific alleles of bcl-2, IL-10, and Fas-L, but not CTLA-4*				
	Allele	OR†	95% CI	
bcl-2	Baseline (195)	1.00	—	—
	191	1.59	1.06, 2.37	0.024
	193	5.61	2.99, 10.53	0.0001‡
	197	1.75	0.72, 4.23	0.215
	199	1.42	0.39, 5.24	0.596
	201	3.15	1.38, 7.17	0.006‡
IL-10	Baseline (125)	1.00	—	—
	121	0.43	0.04, 4.02	0.455
	123	0.63	0.27, 1.50	0.296
	127	2.81	1.78, 4.44	0.0001‡
	129	1.96	1.14, 3.38	0.015
	131	1.28	0.71, 2.29	0.416
	133	1.50	0.96, 2.36	0.077
	135	1.11	0.64, 1.93	0.716
	137	3.73	0.98, 14.22	0.054
Fas-L	Baseline (241)	1.00	—	—
	233	5.77	0.63, 52.9	0.121
	235	2.03	0.44, 9.33	0.365
	237	4.23	0.76, 23.67	0.100
	239	1.69	1.23, 2.33	0.001‡
	243	0.90	0.59, 1.38	0.628
CTLA-4	Baseline (88)	1.00	—	—
	94	0.71	0.11, 4.67	0.724
	102	3.18	0.61, 16.65	0.170
	104	1.12	0.62, 2.02	0.704
	106	0.72	0.50, 1.03	0.070
	108	1.85	0.88, 3.90	0.106
	110	3.06	0.97, 9.67	0.056
	112	1.28	0.20, 8.31	0.793
	114	5.57	0.60, 51.41	0.129
	118	1.19	0.23, 6.16	0.829
	122	2.66	0.24, 29.93	0.428

*2n = number of chromosomes scored to determine allele frequencies.
†OR = odds ratio.
‡P < 0.05.

TABLE 5-continued

Association between SLE and specific alleles of bcl-2, IL-10, and Fas-L, but not CTLA-4*			
Allele	OR†	95% CI	
124	1.32	0.38, 4.64	0.662
126	0.61	0.15, 2.51	0.492
128	1.18	0.16, 8.62	0.865

*Each allele of the IL-10, bcl-2, Fas-L and CTLA-4 loci was compared with a baseline allele of the corresponding locus. The most common allele in the control group for a given locus was chosen as the baseline allele. Shown are the odds ratios (OR) and the 95% confidence intervals (CI) of the association between systemic lupus erythematosus (SLE) and various alleles of the 4 loci compared with baseline.

†Wald χ^2 test, testing H_0 : OR = 1 for each allele, compared with baseline. ‡Significant at the 0.05 level after Bonferroni adjustment, to control the Type I error rate across multiple comparisons within a locus.

TABLE 6

Tests for interaction between loci*			
Locus 1	Locus 2	χ^2	P
IL-10 (127)	Fas-L (239)	0.8†	0.37
IL-10 (127)	bcl-2 (193)	8.11	0.004
Fas-L (239)	bcl-2 (193)	NA‡	—

*The likelihood ratio χ^2 test was used to determine whether each interaction significantly improved the fit compared with a model including only the component main effects of the 2 loci. There were not enough cases and controls carrying high-risk alleles at both bcl-2 and Fas-L to permit estimation of an interaction between these loci.

†Likelihood ratio χ^2 for H_0 : no interaction effect.

‡Not applicable: insufficient data to calculate likelihood ratio χ^2 .

Taken together, the data presented show a novel association between 3 genes involved in apoptosis, bcl-2, Fas-L, and IL-10, and SLE. CTLA-4 did not exhibit an association with SLE. Furthermore, surprisingly, we have demonstrated a synergistic effect between the susceptibility allele 193 of the bcl-2 gene and the susceptibility allele 127 of the IL-10 gene in determining disease susceptibility.

TABLE 7

Synergistic effect of IL-10 and bcl-2 loci on SLE*					
		Sample Size		OR	95% CI
IL-10	bcl-2	Controls	Cases		
x/x	y/y	161	87	1.00	—
127/x, 127/127	y/y	31	27	1.61	0.90, 2.87
x/x	193/y	18	20	2.06	1.03, 4.09
127/x, 127/127	193/y	1	22	40.71	5.40, 307.2

*The odds ratios (ORs) shown are based on the parametric estimates in a logistic model. x indicates any allele other than 127 for IL-10; y indicates any allele other than 193 for bcl-2. SLE = systemic lupus erythematosus; 95% CI = 95% confidence interval.

In a case-control study investigating associations between a disease and one or more genes, there is the potential for bias in odds ratio estimates due to ethnic confounding, commonly called population stratification. Depending on the relationship between an ethnic confounder and the disease, the gene-disease odds ratio may either be positively or negatively biased. In practice, it is impossible to determine the direction of bias unless the confounding variable(s) can be directly measured and controlled for in the analysis. In this study, we minimized the problem by obtaining both cases and controls from the same ethnic group (Mexican Americans), with the additional requirement that the mater-

nal and paternal grandparents of both cases and controls must have been born in Mexico.

The markers we used are short tandem repeat sequences located in the noncoding regions of their respective genes. We relied on principles of linkage disequilibrium in our tests of association and in the corresponding inference that the genes as a whole might play a role in the SLE disease process. Linkage disequilibrium is valid over small genetic distances (within 1 or 2 centimorgans), which obviously covers the intragenic ranges of the genes in the study. In the future it is likely to be found that such sequences are functionally relevant to the expression and biologic properties of these gene products.

Whereas the IL-10 and the bcl-2 genes are both located on chromosome 1 in the mouse, in the human, they reside on separate chromosomes; IL-10 is on chromosome 1q31-q32, and the bcl-2 gene is on 18q21. Therefore, these genes are not in linkage disequilibrium and the appearance of IL-10 susceptibility alleles together with bcl-2 susceptibility alleles in SLE patients represents a true synergism.

EXAMPLE 6

Application of the Test to Other Ethnic Groups

The identification of disease-associated alleles for SLE in a Mexican American population is a clear indication that they will be present in other ethnic groups. However, the specific disease-associated allele may differ. For example the Caucasian and Mexican American population share 80–90% similar genetic background. It is likely that they will share disease-associated alleles. However, other ethnic groups may have different disease-associated alleles. Therefore, the test for genetic predisposition in other ethnic groups would be as follows:

A test group and a control group is identified. A PCR is performed on each of the apoptotic genes, bcl-1, IL-10, Fas-L, and CTLA-4 using the primers as in Example 2. The size of the PCR products is determined. Patients with SLE are compared to a control group to determine the disease-associated allele (by size or sequence). The test involves identifying the presence of that allele for at least two and up to four of the apoptotic genes.

Turner et al (Eur J Immunogenet 1997;24:1–8) identified a single basepair polymorphism at –1082 in the promoter region of the human IL-10 gene which constitutes a G-to-A substitution. Production of IL-10 following concanavalin A stimulation of peripheral blood lymphocytes from individuals carrying a G at position –1082 was significantly increased compared with those with an A at that position. The IL-10 dinucleotide marker used in the present study is located within 50 basepairs of the –1082 G/A polymorphism. It is likely, therefore, that there is linkage disequilibrium between the 2 polymorphisms. Without wishing to be bound by the hypothesis, it is likely that these polymorphisms directly affect transcription factor binding and rates of transcription.

EXAMPLE 7

Application to Other Autoimmune Diseases

Our data on the interaction between bcl-2 and IL-10 underscore the importance of genes that regulate apoptosis in autoimmunity. Transgenic mice that over-express bcl-2 in B lymphocytes exhibit polyclonal expansion and extended survival in vitro. After a few months, these mice develop autoimmune syndromes resembling SLE, including the

11

appearance of antihistone and anti-Sm autoantibodies and immune complex-mediated nephritis. Recent studies in SLE patients suggest that bcl-2 expression is elevated in both B and T lymphocytes.

Therefore, it can be easily envisioned that a comparable test for other autoimmune diseases would follow easily from the above test for SLE. Different autoimmune diseases share susceptibility regions on the chromosomes. Therefore, a test for particularly Thyroid autoimmunity syndromes such as Graves disease, insulin dependent diabetes mellitus, inflammatory bowel disease, rheumatoid arthritis and other arthritides would be apparent from the SLE test.

The test for other autoimmune diseases would be as follows: A test group and a control group is identified. A PCR is performed on each of the apoptotic genes, bcl-1, IL-10, Fas-L, and CTLA-4 using the primers as in Example 2. The size of the PCR products is determined. Patients with SLE are compared to a control group to determine the disease-associated allele (by size or sequence). The test involves identifying the presence of that allele for at least two and up to four of the apoptotic genes. Alternatively different primers could be used for the PCR.

In addition, the test for apoptotic susceptibility loci could be administered with other tests for autoimmunities to get a more definitive diagnosis or test for predisposition. The disease-associated allele would again have to be identified.

Regarding IL-10, elevated levels of this cytokine are found in SLE patients. In addition, IL-10 prevents the spontaneous death of human splenic B cells in vitro, an effect that is abolished by neutralizing anti-IL-10 antibody. IL-10 inhibits apoptotic cell death in human T cells starved of IL-2 and promotes the in vitro survival of T lymphocytes from patients with infectious mononucleosis that, otherwise, are destined to die by apoptosis. These findings are of

12

importance because the continuous administration of IL-10 to lupus-prone (NZB×NZW)F₁ mice accelerated, and neutralizing anti-IL-10 antibody delayed the onset of autoimmunity. Furthermore, the protective effect of IL-10 against B cell death is associated with an increased expression of bcl-2. Our data on the synergistic effect of IL-10 and bcl-2 in human SLE provides a genetic basis for these observations. The data are consistent with the notion that the maintenance of a high-level anti-apoptotic state in lymphocytes contributes to the pathogenesis of SLE by sustaining the rate of production of autoreactive antibodies.

The distal portion of human chromosome 1 (q41-q42) has been recently shown to contain an SLE susceptibility gene. Although the IL-10 gene resides on the distal portion of human chromosome 1, its exact location is proximal to the q41 region and, therefore, IL-10, but not the q41-q42 region, is closer to the chromosomal interval previously mapped in lupus-prone mice. The recent identification of the q41-q42 susceptibility region on chromosome 1, together with our data, identifies a presently unknown, SLE susceptibility gene.

Conclusion

In summary, given the anti-apoptotic nature of bcl-2 and under certain conditions, IL-10, our data further support the notion that inappropriate elimination of autoreactive lymphocytes is an important event in the development of SLE and other autoimmunities. We demonstrate here for the first time that a specific combination of 2 distinct genes that regulate apoptosis identifies a human predisposition to an autoimmune disease. In addition we provide a method for determining genetic predisposition to systemic lupus erythematosus and other autoimmune diseases by genotyping IL-10, bcl-2, Fas ligand and other apoptotic genes.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 8

<210> SEQ ID NO 1
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer to IL-10 microsatellites.

<400> SEQUENCE: 1

gcaaacactcc tcgtcgcaac

20

<210> SEQ ID NO 2
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: PCR primer to IL-10 microsatellite

<400> SEQUENCE: 2

cctcccaaag aagccttagt ag

22

<210> SEQ ID NO 3
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: PCR primer to bcl-2 microsatellite.

<400> SEQUENCE: 3

cgtgtacaca ctctcataca cggct

25

<210> SEQ ID NO 4

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: PCR primer to bcl-2 microsatellite.

<400> SEQUENCE: 4

gggaggggtgc gccatgaaaa

20

<210> SEQ ID NO 5

<211> LENGTH: 25

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: PCR primer to Fas-L microsatellite.

<400> SEQUENCE: 5

cacttctaaa tgcatactct gagcc

25

<210> SEQ ID NO 6

<211> LENGTH: 28

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: PCR primer to Fas-L microsatellite.

<400> SEQUENCE: 6

tgtcaggaag cattcaaat ctgacca

28

<210> SEQ ID NO 7

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: PCR primer to CTLA-4 microsatellite.

<300> PUBLICATION INFORMATION:

<301> AUTHORS: Polymeropoulos, et al.

<303> JOURNAL: Nucleic Acids Research

<304> VOLUME: 19

<305> ISSUE: 1991

<306> PAGES: 4018

<400> SEQUENCE: 7

gccagtgatg ctaaagggtg

20

<210> SEQ ID NO 8

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: PCR primer to CTLA-4 microsatellite.

<300> PUBLICATION INFORMATION:

<301> AUTHORS: Polymeropoulos, et al.

<303> JOURNAL: Nucleic Acids Research

<304> VOLUME: 19

<305> ISSUE: 1991

<306> PAGES: 4018

<400> SEQUENCE: 8

aacatacgtg gctctatgca

20

15

What is claimed is:

1. A method for determining predisposition to an autoimmune disease in a patient, comprising:

a) obtaining a sample containing genetic material from said patient; and

b) determining whether alleles associated with susceptibility to said autoimmune disease are present in both IL-10 and bcl-2 loci in said sample, wherein the presence of both said alleles indicates that said patient has a predisposition to said autoimmune disease.

2. The method of claim 1, wherein the determining step comprises amplification of said genetic material.

3. The method of claim 2, wherein the amplification makes use of a primer specific for said allele associated with susceptibility to said autoimmune disease.

4. The method of claim 1, wherein the determining step comprises hybridization with a probe specific for said allele associated with susceptibility to said autoimmune disease.

5. The method of claim 1, wherein the IL-10 gene is amplified with primers comprising the sequences of SEQ ID NO:1 and SEQ ID NO:2.

16

6. The method of claim 2 wherein a bcl-2 disease-specific allele is amplified with primers comprising the sequences of SEQ ID NO:3 and SEQ ID NO:4.

7. The method of claim 1 wherein the allele associated with susceptibility to said autoimmune disease is identified by its size.

8. The method of claim 1 wherein the alleles associated with susceptibility to said autoimmune disease are IL-10 (127) and bcl-2(193) and wherein presence of both alleles indicates a greater likelihood of predisposition to said autoimmune disease than presence of either allele alone.

9. The method of claim 1 wherein the autoimmune disease is selected from the group consisting of; systemic lupus erythematosus, thyroid autoimmunity syndromes, insulin dependent diabetes mellitus, inflammatory bowel disease, rheumatoid arthritis and other arthritides.

10. The method of claim 9 wherein the disease is systemic lupus erythematosus.

* * * * *

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant	:	Eaton, et al.
Appl. No.	:	10/063,557
Filed	:	May 2, 2002
For	:	SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME
Examiner	:	David J. Blanchard
Group Art Unit	:	1642

DECLARATION OF J. CHRISTOPHER GRIMALDI UNDER 37 C.F.R. § 1.132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

I, J. Christopher Grimaldi, declare and say as follows:

1. I am a Senior Research Associate in the Molecular Biology Department of Genentech, Inc., South San Francisco, CA 94080.

2. I joined Genentech in January of 1999. From 1999 to 2003, I directed the Cloning Laboratory in the Molecular Biology Department. During this time I directed or performed numerous molecular biology techniques including qualitative Polymerase Chain Reaction (PCR) analyses. I am currently involved in, among other projects, the isolation of genes coding for membrane associated proteins which can be used as targets for antibody therapeutics against cancer. In connection with the above-identified patent application, I personally performed or directed the semi-quantitative PCR analyses in the assay entitled "Tumor Versus Normal Differential Tissue Expression Distribution" which is described in EXAMPLE 18 in the specification that were used to identify differences in gene expression between tumor tissue and their normal counterparts.

3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).

4. In differential gene expression studies, one looks for genes whose expression levels differ significantly under different conditions, for example, in normal versus diseased tissue.

Appl. No. : 10/063,557
Filed : May 2, 2002

Chromosomal aberrations, such as gene amplification, and chromosomal translocations are important markers of specific types of cancer and lead to the aberrant expression of specific genes and their encoded polypeptides, including over-expression and under-expression. For example, gene amplification is a process in which specific regions of a chromosome are duplicated, thus creating multiple copies of certain genes that normally exist as a single copy. Gene under-expression can occur when a gene is not transcribed into mRNA. In addition, chromosomal translocations occur when two different chromosomes break and are rejoined to each other chromosome resulting in a chimeric chromosome which displays a different expression pattern relative to the parent chromosomes. Amplification of certain genes such as Her2/Neu [Singleton *et al.*, Pathol. Annu., 27Pt1:165-190], or chromosomal translocations such as t(5;14), [Grimaldi *et al.*, Blood, 73(8):2081-2085(1989); Meeker *et al.*, Blood, 76(2):285-289(1990)] give cancer cells a growth or survival advantage relative to normal cells, and might also provide a mechanism of tumor cell resistance to chemotherapy or radiotherapy. When the chromosomal aberration results in the aberrant expression of a mRNA and the corresponding gene product (the polypeptide), as it does in the aforementioned cases, the gene product is a promising target for cancer therapy, for example, by the therapeutic antibody approach.

5. Comparison of gene expression levels in normal versus diseased tissue has important implications both diagnostically and therapeutically. For example, those who work in this field are well aware that in the vast majority of cases, when a gene is over-expressed, as evidenced by an increased production of mRNA, the gene product or polypeptide will also be over-expressed. It is unlikely that one identifies increased mRNA expression without associated increased protein expression. This same principle applies to gene under-expression. When a gene is under-expressed, the gene product is also likely to be under-expressed. Stated in another way, two cell samples which have differing mRNA concentrations for a specific gene are expected to have correspondingly different concentration of protein for that gene. Techniques used to detect mRNA, such as Northern Blotting, Differential Display, *in situ* hybridization, quantitative PCR, Taqman, and more recently Microarray technology all rely on the dogma that a change in mRNA will represent a similar change in protein. If this dogma did not hold true then these techniques would have little value and not be so widely used. The use of mRNA quantitation techniques have identified a seemingly endless number of genes which are differentially expressed in various tissues and these genes have subsequently been shown to have correspondingly similar changes in their protein levels. Thus, the detection of increased mRNA expression is expected to result in increased polypeptide expression, and the detection of decreased mRNA expression is expected to result in decreased polypeptide expression. The detection of increased or decreased polypeptide expression can be used for cancer diagnosis and treatment.

6. However, even in the rare case where the protein expression does not correlate with the mRNA expression, this still provides significant information useful for cancer diagnosis and treatment. For example, if over- or under-expression of a gene product does not correlate with over- or under-expression of mRNA in certain tumor types but does so in others, then identification of both gene expression and protein expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over- or

Appl. No. : 10/063,557
Filed : May 2, 2002

under-expression of the gene product in the presence of a particular over- or under-expression of mRNA is crucial information for the practicing clinician. For example, if a gene is over-expressed but the corresponding gene product is not significantly over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: _____

J. Christopher Grimaldi

Date: _____

8/10/2004

S:\DOCS\AOK\ADK-5479.DOC
071904

EXHIBIT A

J. Christopher Grimaldi

1434-36th Ave.
San Francisco, CA 94122
(415) 681-1639 (Home)

EDUCATION

University of California, Berkeley
Bachelor of Arts in Molecular Biology, 1984

EMPLOYMENT EXPERIENCE

SRA

Genentech Inc., South San Francisco; 1/99 to present

Previously, was responsible to direct and manage the Cloning Lab. Currently focused on isolating cancer specific genes for the Tumor Antigen (TAP), and Secreted Tumor Protein (STOP) projects for the Oncology Department as well as Immunologically relevant genes for the Immunology Department. Directed a lab of 6 scientists focused on a company-wide team effort to identify and isolate secreted proteins for potential therapeutic use (SPDI). For the SPDI project my duties were, among other things, the critically important coordination of the cloning of thousands of putative genes, by developing a smooth process of communication between the Bioinformatics, Cloning, Sequencing, and Legal teams. Collaborated with several groups to discover novel genes through the Curagen project, a unique differential display methodology. Interacted extensively with the Legal team providing essential data needed for filing patents on novel genes discovered through the SPDI, TAP and Curagen projects. My group has developed, implemented and patented high throughput cloning methodologies that have proven to be essential for the isolation of hundreds of novel genes for the SPDI, TAP and Curagen projects as well as dozens of other smaller projects.

Scientist

DNAX Research Institute, Palo Alto; 9/91 to 1/99

Involved in multiple projects aimed at understanding novel genes discovered through bioinformatics studies and functional assays. Developed and patented a method for the specific depletion of eosinophils in vivo using monoclonal antibodies. Developed and implemented essential technical methodologies and provided strategic direction in the areas of expression, cloning, protein purification, general molecular biology, and monoclonal antibody production. Trained and supervised numerous technical staff.

Facilities Manager

Corixa, Redwood City; 5/89 - 7/91.

Directed plant-related activities, which included expansion planning, maintenance, safety, purchasing, inventory control, shipping and receiving, and laboratory management. Designed and implemented the safety program. Also served as liaison to regulatory agencies at the local, state and federal level. Was in charge of property leases, leasehold improvements, etc. Negotiated vendor contracts and directed the purchasing department. Trained and supervised personnel to carry out the above-mentioned duties.

SRA University of California, San Francisco
Cancer Research Institute; 2/87-4/89.

Was responsible for numerous cloning projects including: studies of somatic hypermutation, studies of AIDS-associated lymphomas, and cloning of t(5;14), t(11;14), and t(8;14) translocations. Focused on the activation of hemopoietic growth factors involved in the t(5;14) translocation in leukemia patients..

Research Berlex Biosciences, South San Francisco; 7/85-2/87.
Technician

Worked on a subunit porcine vaccine directed against *Mycoplasma hyopneumoniae*. Was responsible for generating genomic libraries, screening with degenerate oligonucleotides, and characterizing and expressing clones in *E. coli*. Also constructed a general purpose expression vector for use by other scientific teams.

PUBLICATIONS

1. Hilary F. Clark, et al. "The Secreted Protein Discovery Initiative (SPDI), a Large-scale Effort to Identify Novel Human Secreted and Transmembrane Proteins: a bioinformatics assessment." *Genome Res.* Vol 13(10), 2265-2270, 2003
2. Sean H. Adams, Clarissa Chui, Sarah L. Schilbach, Xing Xian Yu, Audrey D. Goddard, J. Christopher Grimaldi, James Lee, Patrick Dowd, David A. Lewin, & Steven Colman. "BFTT, a Unique Acyl-CoA Thioesterase Induced in Thermogenic Brown Adipose Tissue: Cloning, organization of the human gene and assessment of a potential link to obesity" *Biochemical Journal*, Vol 360, 135-142, 2001
3. Szeto W, Jiang W, Tice DA, Rubinfeld B, Hollingshead PG, Fong SE, Dugger DL, Pham T, Yansura D, Wong TA, Grimaldi JC, Corpuz RT, Singh JS, Frantz GD, , Devaux B, Crowley CW, Schwall RH, Eberhard DA, Rastelli L, Polakis P, and Pennica D. "Overexpression of the Retenoic Acid-Responsive Gene *Stra6* in Human Cancers and its Synergistic Activation by Wnt-1 and Retinoic Acid." *Cancer Research* Vol. 61(10), 4197-4205, 2001
4. Jeanne Kahn, Fuad Mehraban, Gladdys Ingle, Xiaohua Xin, Juliet E. Bryant, Gordon Vehar, Jill Schoenfeld, J. Christopher Grimaldi (incorrectly named as "Grimaldi, CJ"), Franklin Peale, Aparna Draksharapu, David A. Lewin, and Mary E. Gerritsen. "Gene Expression Profiling in an in Vitro Model of Angiogenesis." *American Journal of Pathology* Vol 156(6), 1887-1900, 2000.
5. Grimaldi JC, Yu NX, Grunig G, Seymour BW, Cottrez F, Robinson DS, Hosken N, Ferlin WG, Wu X, Soto H, O'Garra A, Howard MC, Coffman RL. "Depletion of eosinophils in mice through the use of antibodies specific for C-C chemokine receptor 3 (CCR3). *Journal of Leukocyte Biology*; Vol. 65(6), 846-53, 1999
6. Oliver AM, Grimaldi JC, Howard MC, Kearney JF. "Independently ligating CD38 and Fc gammaRIIB relays a dominant negative signal to B cells." *Hybridoma* Vol. 18(2), 113-9, 1999

7. Cockayne DA, Muchamuel T, Grimaldi JC, Muller-Steffner H, Randall TD, Lund FE, Murray R, Schuber F, Howard MC. "Mice deficient for the ecto-nicotinamide adenine dinucleotide glycohydrolase CD38 exhibit altered humoral immune responses." *Blood* Vol. 92(4), 1324-33, 1998
8. Frances E. Lund, Nanette W. Solvason, Michael P. Cooke, Andrew W. Heath, J. Christopher Grimaldi, Troy D. Randall, R. M. E. Parkhouse, Christopher C Goodnow and Maureen C. Howard. "Signaling through murine CD38 is impaired in antigen receptor unresponsive B cells." *European Journal of Immunology*, Vol. 25(5), 1338-1345, 1995
9. M. J. Guimaraes, J. F. Bazan, A. Zolotnik, M. V. Wiles, J. C. Grimaldi, F. Lee, T. McClanahan. "A new approach to the study of haematopoietic development in the yolk sac and embryoid body." *Development*, Vol. 121(10), 3335-3346, 1995
10. J. Christopher Grimaldi, Sriram Balasubramanian, J. Fernando Bazan, Armen Shanafelt, Gerard Zurawski and Maureen Howard. "CD38-mediated protein ribosylation." *Journal of Immunology*, Vol. 155(2), 811-817, 1995
11. Leopoldo Santos-Argumedo, Frances F. Lund, Andrew W. Heath, Nanette Solvason, Wei Wei Wu, J. Christopher Grimaldi, R. M. E. Parkhouse and Maureen Howard. "CD38 unresponsiveness of xid B cells implicates Bruton's tyrosine kinase (btk) as a regulator of CD38 induced signal transduction." *International Immunology*, Vol 7(2), 163-170, 1995
12. Frances Lund, Nanette Solvason, J. Christopher Grimaldi, R. M. E. Parkhouse and Maureen Howard. "Murine CD38: An immunoregulatory ectoenzyme." *Immunology Today*, Vol. 16(10), 469-473, 1995
13. Maureen Howard, J. Christopher Grimaldi, J. Fernando Bazan, Frances E. Lund, Leopoldo Santos-Argumedo, R. M. E. Parkhouse, Timothy F. Walseth, and Hon Cheung Lee. "Formation and Hydrolysis of Cyclic ADP-Ribose Catalyzed by Lymphocyte Antigen CD38." *Science*, Vol. 262, 1056-1059, 1993
14. Nobuyuki Harada, Leopoldo Santos-Argumedo, Ray Chang, J. Christopher Grimaldi, Frances Lund, Camilynn I. Brannan, Neal G. Copeland, Nancy A. Jenkins, Andrew Heath, R. M. E. Parkhouse and Maureen Howard. "Expression Cloning of a cDNA Encoding a Novel Murine B Cell Activation Marker: Homology to Human CD38." *The Journal of Immunology*, Vol. 151, 3111-3118, 1993
15. David J. Rawlings, Douglas C. Saffran, Satoshi Tsukada, David A. Largaespada, J. Christopher Grimaldi, Lucie Cohen Randolph N. Mohr, J. Fernando Bazan, Maureen Howard, Neal G. Copeland, Nancy A. Jenkins, Owen Witte. "Mutation of Unique Region of Bruton's Tyrosine Kinase in Immunodeficient XID Mice." *Science*, Vol. 261, 358-360, 1993
16. J. Christopher Grimaldi, Raul Torres, Christine A. Kozak, Ray Chang, Edward Clark, Maureen Howard, and Debra A. Cockayne. "Genomic Structure and Chromosomal Mapping of the Murine CD40 Gene." *The Journal of Immunology*, Vol 149, 3921-3926, 1992
17. Timothy C. Meeker, Bruce Shiramizu, Lawrence Kaplan, Brian Herndier, Henry Sanchez, J. Christopher Grimaldi, James Baumgartner, Jacob Rachlin, Ellen Feigal, Mark Rosenblum and Michael S. McGrath. "Evidence for Molecular Subtypes of HIV-Associated Lymphoma:

Division into Peripheral Monoclonal, Polyclonal and Central Nervous System Lymphoma." AIDS, Vol. 5, 669-674, 1991

18. Ann Grimaldi and Chris Grimaldi. "Small-Scale Lambda DNA Prep." Contribution to Current Protocols in Molecular Biology, Supplement 5, Winter 1989
19. J. Christopher Grimaldi, Timothy C. Meeker. "The t(5;14) Chromosomal Translocation in a Case of Acute Lymphocytic Leukemia Joins the Interleukin-3 Gene to the Immunoglobulin Heavy Chain Gene." Blood, Vol. 73, 2081-2085, 1989
20. Timothy C. Meeker, J. Christopher Grimaldi, et al. "An Additional Breakpoint Region in the BCL-1 Locus Associated with the t(11;14) (q13;q32) Translocation of B-Lymphocytic Malignancy." Blood, Vol. 74, 1801-1806, 1989
- 21 Timothy C. Meeker, J. Christopher Grimaldi, Robert O'Rourke, et al. "Lack of Detectable Somatic Hypermutation in the V Region of the Ig H Chain Gene of a Human Chronic B Lymphocytic Leukemia." The Journal of Immunology, Vol. 141, 3994-3998, 1988

MANUSCRIPTS IN PREPARATION

1. Sriram Balasubramanian, J. Christopher Grimaldi, J. Fernando Bazan, Gerard Zurawski and Maureen Howard. "Structural and functional characterization of CD38: Identification of active site residues"

PATENTS

1. "Methods for Eosinophil Depletion with Antibody to CCR3 Receptor" (US 6,207,155 B1).
2. "Amplification Based Cloning Method." (US 6,607,899)
3. Ashkenazi et al., "Secreted and Transmembrane Polypeptides and Nucleic Acids Encoding the Same." (this patent covers several hundred genes)
4. "IL-17 Homologous Polypeptides and Therapeutic Uses Thereof"
5. "Method of Diagnosing and Treating Cartilaginous Disorders."

MEMBERSHIPS AND ACTIVITIES

Editor Frontiers in Bioscience

Member DNAX Safety Committee 1991-1999
 Biological Safety Affairs Forum (BSAF) 1990-1991
 Environmental Law Foundation (ELF) 1990-1991

The t(5;14) Chromosomal Translocation in a Case of Acute Lymphocytic Leukemia Joins the Interleukin-3 Gene to the Immunoglobulin Heavy Chain Gene

By J. Christopher Grimaldi and Timothy C. Meeker

Chromosomal translocations have proven to be important markers of the genetic abnormalities central to the pathogenesis of cancer. By cloning chromosomal breakpoints one can identify activated proto-oncogenes. We have studied a case of B-lineage acute lymphocytic leukemia (ALL) that was associated with peripheral blood eosinophilia. The chromosomal translocation t(5;14) (q31;q32) from this sample was cloned and studied at the molecular level. This

translocation joined the immunoglobulin heavy chain joining (Jh) region to the promotor region of the interleukin-3 (IL-3) gene in opposite transcriptional orientations. The data suggest that activation of the IL-3 gene by the enhancer of the immunoglobulin heavy chain gene may play a central role in the pathogenesis of this leukemia and the associated eosinophilia.

© 1989 by Grune & Stratton, Inc.

KARYOTYPIC STUDIES of leukemia and lymphoma have identified frequent nonrandom chromosomal translocations. Some of these translocations juxtapose the immunoglobulin heavy chain (IgH) gene with important

protooncogenes, such as *c-myc* and *bcl-2*.^{1,2} In this way, the IgH gene can activate proto-oncogenes, resulting in disordered gene expression and a step in the development of cancer. The investigation of additional nonrandom translocations into the IgH locus allows us to identify new genes promoting the generation of leukemia and lymphoma.

A distinct subtype of acute lymphocytic leukemia (ALL) has been characterized by B-lineage phenotype, associated eosinophilia in the peripheral blood, and a t(5;14)(q31;q32) chromosomal translocation.^{3,4} This syndrome probably occurs in <1% of all patients with ALL. We hypothesized that the cloning of the translocation characteristic of this leukemia might allow the identification of an important gene on chromosome 5 that plays a role in the evolution of this disease. In this report we demonstrate that the interleukin-3 gene (IL-3) and the IgH gene are joined by this translocation.

MATERIALS AND METHODS

Sample and DNA blots. A bone marrow aspirate from a representative patient with ALL (L1 morphology by French-American-British [FAB] criteria), peripheral eosinophilia (up to 20,000 per microliter with a normal value of <350 per microliter) and a t(5;14)(q31;q32) translocation was studied. Using published methods, genomic DNA was isolated and DNA blots were made.⁵ Briefly, 10 µg of high molecular weight (mol wt) DNA were digested using an appropriate restriction enzyme and electrophoresed on a 0.8% agarose gel. The gel was stained with ethidium bromide, photographed, denatured, neutralized, and transferred to Hybond (Amersham, Arlington Heights, IL). After treatment of the filter with ultraviolet light, hybridization was performed. The filter was washed to a final stringency of 0.2% saturated sodium citrate (SSC) and 0.1% sodium lauryl sulfate (SDS) and exposed to film. The human Jh probe has been previously reported.⁶

Genomic library. The genomic library was made using pub-

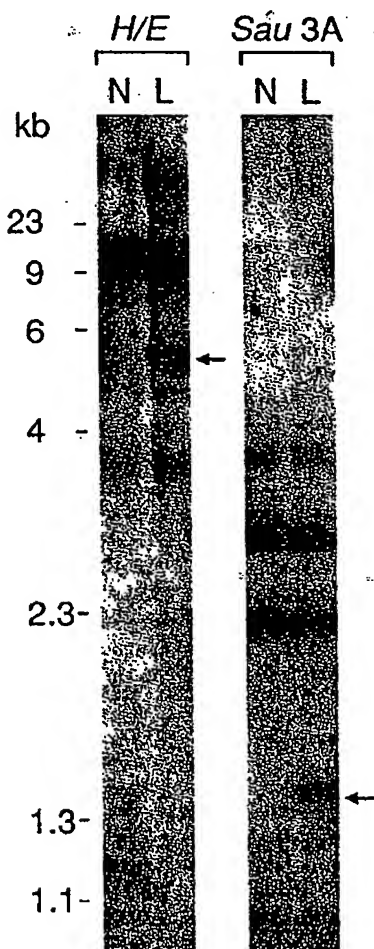


Fig 1. DNA blots of the leukemia sample. The restriction fragment pattern of normal human DNA (N) and the leukemia sample (L) were compared using a human Jh probe. Rearranged bands are indicated by arrows. Sample L exhibits a single rearranged band with both *Hind*III/*Eco*RI and *Sau*3A restriction digests. The rearranged bands are less intense than the other bands because the majority of cells in the sample represent normal bone marrow elements.

From the Division of Hematology/Oncology, Department of Medicine, University of California, San Francisco.

Submitted February 22, 1989; accepted March 8, 1989.

Supported by NIH Grant No. CA01102.

Address reprint requests to Timothy C. Meeker, MD, UCSF/VAMC 111H, 4150 Clement St, San Francisco, CA 94121.

Dr. Grimaldi's current address is Biostan Inc, 440 Chesapeake Dr, Seaport Centre, Redwood City, CA 94063.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. section 1734 solely to indicate this fact.

© 1989 by Grune & Stratton, Inc.
0006-4971/89/7308-0031\$3.00/0

lished methods.⁵ Approximately 100 μ g of high mol wt genomic DNA were partially digested with the *Sau*3A restriction enzyme. Fragments from 9 to 23 kilobases (kb) in size were isolated on a sucrose gradient and ligated into phage EMBL3A (Stratagene, San Diego). Recombinant phage were packaged, plated, and screened as previously reported.⁵

DNA sequencing. Fragments for sequencing were cloned into M13 vectors and sequenced by the chain termination method using Sequenase (United States Biochemical, Cleveland).⁷ All sequence data were derived from both strands.

RESULTS

We studied a bone marrow sample from a patient with ALL and associated peripheral eosinophilia. Karyotypic analysis showed the characteristic t(5;14)(q31;q32) translocation. These features define a distinctive subtype of ALL.^{3,4} The leukemic cells were analyzed for cell surface phenotype by immunofluorescence. They were positive for B1 (CD20), B4 (CD19), cALLA (CD10), HLA-DR, and terminal deoxynucleotidyl transferase (Tdt), but negative for surface immunoglobulin. This phenotypic profile describes an immature cell from the B-lymphocytic lineage.⁸

The leukemia DNA was analyzed by Southern blotting for rearrangements of the IgH gene. Using a human immunoglobulin Jh probe, a single rearranged band was detected by *Eco*RI, *Hind*III, *Sst*I, *Sau*3A, and *Eco*RI plus *Hind*III restriction digests, suggesting rearrangement of one allele (Fig 1). The immunoglobulin Jh region from the other allele was presumably either deleted or in the germline configuration.

We hypothesized that the t(5;14)(q31;q32) juxtaposed a

growth-promoting gene on chromosome 5 with the immunoglobulin Jh region on chromosome 14. Therefore, a genomic library was made from the leukemic sample and screened with a Jh probe. Fifteen distinct positive clones were isolated and screened for the presence of the rearranged *Sau*3A fragment that was detected by DNA blotting. By this analysis, five clones appeared to represent the rearranged allele identified by DNA blots. One of these clones (clone no. 4) was chosen for further study and a detailed restriction map was generated. The *Eco*RI, *Hind*III/*Eco*RI, and *Sst*I fragments from clone no. 4 that hybridized to the human Jh probe were also identical in size to the rearranged fragments from the leukemia sample, confirming that clone no. 4 represented the rearranged leukemic allele.

Phage clone no. 4 contained 3.7 kb of unknown origin joined to the IgH gene in the region of Jh4 (Fig 2). The IgH gene from Jh4 to the Cmu region appeared to be in germline configuration. Previously, the gene encoding hematopoietic growth factor IL-3 had been mapped to chromosome 5q31 so it was suspected that clone no. 4 might contain part of this gene.⁹⁻¹² When the restriction map of human IL-3 and clone no. 4 were compared, they were identical for more than 3 kb (Fig 2).

We confirmed the juxtaposition of the IL-3 gene and the IgH gene by nucleic acid sequencing of the subcloned *Bst*EII/*Hpa*I fragment (Fig 2). The sequence of this fragment showed no disruption of the protein coding region or the messenger RNA of the IL-3 gene. The break in the IL-3 gene occurred in the promoter region, 452 base pairs (bp) upstream of the transcriptional start site (position 64, Fig

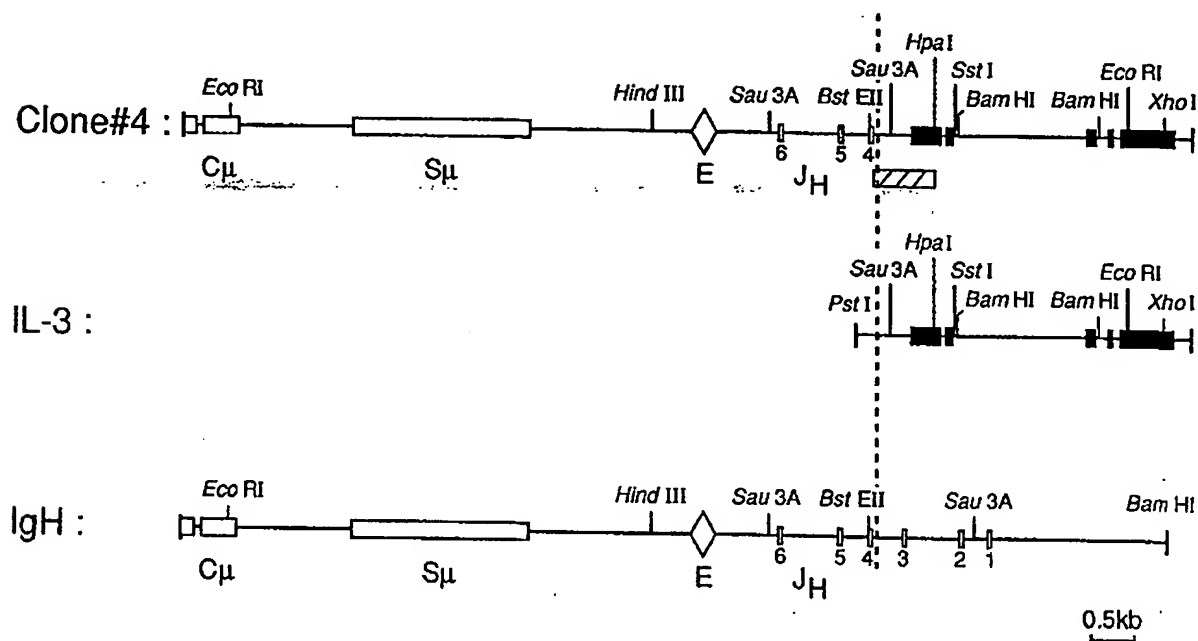


Fig 2. Breakpoint region: t(5;14)(q31;q32). Comparative mapping of phage clone no. 4, the germline IgH region, and the germline IL-3 gene.^{20,29} The map of clone no. 4 is identical to that of IgH until it diverges in the region of Jh4 (at the dashed line), after which it is identical to the map of IL-3. The two genes are positioned in a head-to-head orientation. The Ig μ chain constant region (*C μ*), switch region (*S μ*), enhancer (E), and Jh segments are indicated (open symbols). The five exons (dark boxes) and four introns of the IL-3 gene are shown. The hatched box indicates the sequenced region.

3A). The break in the IgH gene occurred 2 bp upstream of the Jh4 region. Between the two breaks, 25 bp of uncertain origin (putative N sequence) were inserted.^{13,14} No sequences homologous to the immunoglobulin heptamer and nonamer could be identified in the IL-3 sequence (Fig 3B). Therefore, nucleic acid sequencing confirmed the juxtaposition of the IL-3 gene and the IgH gene. The sequence data clearly showed that the genes were positioned in opposite transcriptional orientations (head-to-head).

Available data also allowed us to determine the normal positions of the IL-3 gene and the GM-CSF gene in relation to the centromere of chromosome 5 (Fig 4). The IgH gene is known to be positioned with the variable regions toward the telomere on chromosome 14q.^{2,15} It has also been shown that

GM-CSF maps within 9 kb of IL-3 in the same transcriptional orientation.¹⁶ Using this information and assuming a simple translocation event in our sample, we can conclude that the IL-3 gene is normally more centromeric, and the GM-CSF gene more telomeric on chromosome 5q (Fig 4). Furthermore, both are transcribed with their 5' ends toward the centromere.

DISCUSSION

In this report we have cloned a unique chromosomal translocation that appears to be a consistent feature of a rare, yet distinct, clinical form of acute leukemia. This translocation joined the promoter of the IL-3 gene to the IgH gene. Except for the altered promoter, the IL-3 gene appeared

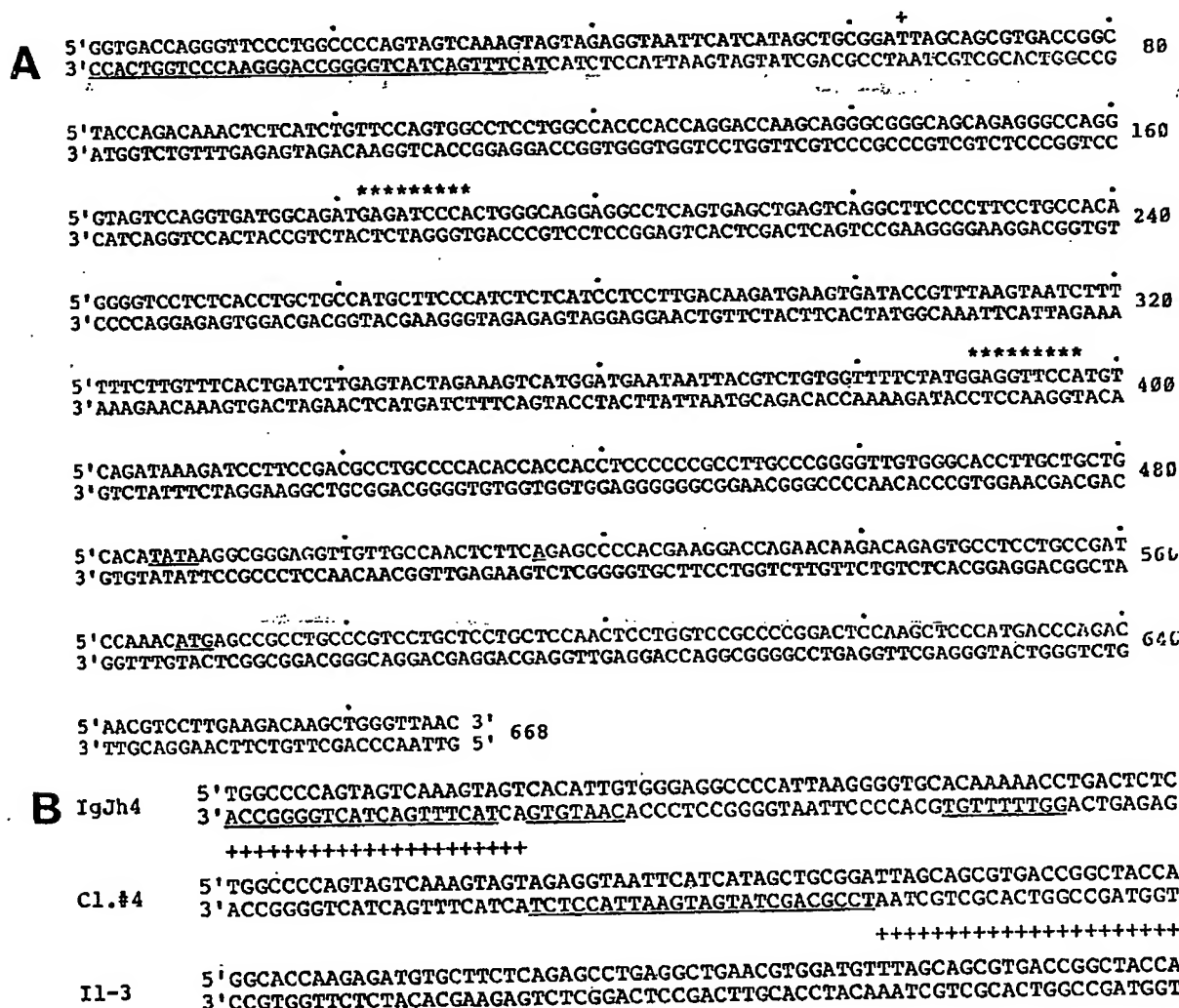


Fig 3. Sequence of t(5;14)(q31;q32) breakpoint region. (A) Nucleotide sequence of the *BstEII/HpaI* fragment indicated on Fig 2. Nucleotides 1 to 36 represent the Jh4 coding region underlined on the coding strand.⁸ Nucleotides 39 to 63 are a putative N region. The sequence from position 64 to 688 is that of the germline IL-3 gene.²⁰ The IL-3 TATA box (485), transcription start (515), and initiation methionine (567) are underlined. Two proposed regulatory sequences in the promoter are marked by asterisks (positions 182 and 389). (B) Comparative sequence of the t(5;14)(q31;q32) breakpoint region. The IgJh4 region is shown with its coding region, heptamer, and nonamer underlined. Clone no. 4 is shown with putative N region sequences underlined. The IL-3 sequence is also shown. A plus sign (+) denotes the identical nucleotide between sequences. No heptamer or nonamer is identified in the IL-3 sequence.

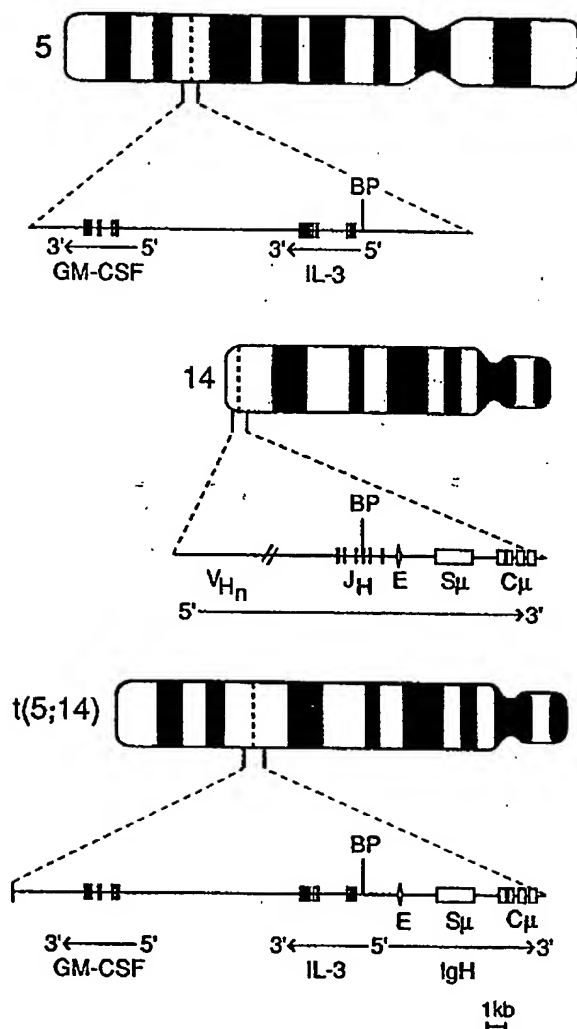


Fig 4. Diagram of the translocation. The normal chromosome 5q31 is shown with the GM-CSF gene telomeric to the IL-3 gene in the transcriptional orientation shown. On normal chromosome 14q32 the V_{Hn} regions are telomeric. The t(5;14)(q31;q32) translocation results in the head-to-head orientation of these genes. Symbols are defined in Fig 2. BP, breakpoint position.

intact as no deletions, insertions, or point mutations were detected by restriction mapping of the entire gene and sequencing of part of the gene. The IgH gene has been truncated at the J_H4 region, which places the immunoglobulin enhancer within 2.5 kb of the IL-3 gene.^{17,18} This leads to the hypothesis that the enhancer is increasing transcription of a structurally normal IL-3 gene. The same mechanism is important for activation of the *c-myc* gene in some cases of Burkitt's lymphoma.¹⁹ An alternate hypothesis is that the elimination of an upstream IL-3 promoter element is crucial to the activation of the IL-3 gene.

The proposed activation of the IL-3 gene suggests that an autocrine loop is important for the pathogenesis of this leukemia.²⁰ Over-expression of the IL-3 gene coupled with

the presence of the IL-3 receptor in these cells could account for a strong stimulus for proliferation. In this regard, there are data indicating that immature B-lineage lymphocytes and B-lineage leukemias may express the IL-3 receptor.^{21,22}

An additional feature of this type of leukemia is the dramatic eosinophilia, consisting of mature forms. It has been hypothesized that the eosinophils do not arise from the malignant clone, but are stimulated by the tumor.^{23,24} Because of the known effect of IL-3 on eosinophil differentiation, secretion of high levels of IL-3 by leukemic cells might have a role in the eosinophilia in this type of leukemia.¹²

The data suggest that the recombination mechanism that is active in the IgH gene during normal differentiation has a role in this translocation.^{13,14} This is supported by the breakpoint location at the 5' end of J_H4 and the presence of putative N-region sequences. On the other hand, no recombination signal sequence (heptamer and nonamer) was found in this region on chromosome 5, suggesting that additional factors also played a role. Further studies will elucidate the mechanism of this and other translocations.

In the leukemia we studied, it is possible that the immunoglobulin enhancer also activates the GM-CSF gene, since this gene is probably positioned only 14 kb away (Fig 4). This is known to be within the range of enhancer activation.²⁵ The interleukin-5 (IL-5) gene maps to chromosome 5q31.²⁶ Deregulation of the IL-5 gene by this translocation would act synergistically with IL-3 in the stimulation of eosinophil proliferation and differentiation.²⁷ These and other questions will be answered by the study of more patient samples. We plan to determine whether the t(5;14)(q31;q32) translocation is capable of activating multiple lymphokines simultaneously and whether they cooperate in the generation of this leukemia.

REFERENCES

1. Klein G, Klein E: Evolution of tumours and the impact of molecular oncology. *Nature* 315:190, 1985
2. Showe L, Croce C: The role of chromosomal translocations in B- and T-cell neoplasia. *Annu Rev Immunol* 5:253, 1987
3. Hogan T, Koss W, Murgo A, Amato R, Fontana J, Van Scoy F: Acute lymphoblastic leukemia with chromosomal 5;14 translocation and hypereosinophilia: Case report and literature review. *J Clin Oncol* 5:382, 1987
4. Tono-oka T, Sato Y, Matsumoto T, Ueno N, Ohkawa M, Shikano T, Takeda T: Hypereosinophilic syndrome in acute lymphoblastic leukemia with a chromosome translocation t(5q;14q). *Med Pediatr Oncol* 12:33, 1984
5. Meeker T, Grimaldi JC, O'Rourke R, Loeb J, Juliusson G, Einhorn S: Lack of detectable somatic hypermutation in the V region of the IgH gene of a human chronic B-lymphocytic leukemia. *J Immunol* 141:3394, 1988
6. Ravetch J, Siebenlist U, Korsmeyer S, Waldmann T, Leder P: Structure of the human immunoglobulin μ locus: Characterization of embryonic and rearranged J and D genes. *Cell* 27:583, 1981
7. Norrander U, Kempe T, Messing J: Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 26:101, 1983
8. Foon K, Todd R: Immunologic classification of leukemia and lymphoma. *Blood* 68:1, 1986
9. LeBeau M, Epstein N, O'Brien SJ, Nienhuis AW, Yang Y-C, Clark S, Rowley J: The interleukin-3 gene is located on human

chromosome 5 and is deleted in myeloid leukemias with a deletion of 5q. *Proc Natl Acad Sci USA* 84:5913, 1987

10. LeBeau M, Chandrasekharappi S, Lemons R, Schwartz J, Larson R, Arai N, Westbrook C: Molecular and cytogenetic analysis of chromosome 5 abnormalities in myeloid disorders, in cancer cells, in *Proceedings of Molecular Diagnostics of Human Cancer*. Cold Spring Harbor Laboratory, NY, 1989 (in press)

11. Ihle J, Weinstein Y: Immunological regulation of hematopoietic/lymphoid stem cell differentiation by interleukin-3. *Adv Immunol* 39:1, 1986

12. Clark S, Kamen R: The human hematopoietic colony-stimulating factors. *Science* 236:1229, 1987

13. Bakhshi A, Wright J, Graninger W, Seto M, Owens J, Cossman J, Jensen J, Goldman P, Korsmeyer S: Mechanism of the t(14;18) chromosomal translocation: Structural analysis of both derivative 14 and 18 reciprocal partners. *Proc Natl Acad Sci USA* 84:2396, 1987

14. Tsujimoto Y, Louie E, Bashir M, Croce C: The reciprocal partners of both the t(14;18) and the t(11;14) translocations involved in B-cell neoplasms are rearranged by the same mechanism. *Oncogene* 2:347, 1988

15. Erikson J, Finan J, Nowell P, Croce C: Translocation of immunoglobulin VH genes in Burkitt lymphoma. *Proc Natl Acad Sci USA* 80:810, 1982

16. Yang Y-C, Kovacic S, Kriz R, Wolf S, Clark S, Wellems T, Nienhuis A, Epstein N: The human genes for GM-CSF and IL-3 are closely linked in tandem on chromosome 5. *Blood* 71:958, 1988

17. Gillies S, Morrison S, Oi V, Tonegawa S: A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33:717, 1983

18. Banerji J, Olson L, Schaffner W: A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* 33:729, 1983

19. Hayday A, Gillies S, Saito H, Wood C, Wiman K, Hayward

W, Tonegawa S: Activation of a translocated human *c-myc* gene by an enhancer in the immunoglobulin heavy-chain locus. *Nature* 307:334, 1984

20. Sporn M, Roberts A: Autocrine growth factors and cancer. *Nature* 313:745, 1985

21. Palacios R, Steinmetz M: IL-3-dependent mouse clones that express B-220 surface antigen, contain Ig genes in germ line configuration, and generate B lymphocytes in vivo. *Cell* 41:727, 1985

22. Uckun F, Gesner T, Song C, Myers D, Mufson A: Leukemic B-cell precursors express functional receptors for human interleukin-3. *Blood* 73:533, 1989

23. Spitzer G, Garson O: Lymphoblastic leukemia with marked eosinophilia: A report of two cases. *Blood* 42:377, 1973

24. Catovsky D, Bernasconi C, Verkonck P, Postma A, Howss J, Berg A, Rees J, Castelli G, Morra E, Galton D: The association of eosinophilia with lymphoblastic leukemia or lymphoma: A study of seven patients. *Br J Haematol* 45:523, 1980

25. Wang X-F, Calame K: The endogenous immunoglobulin heavy chain enhancer can activate tandem Vh promoters separated by a large distance. *Cell* 43:659, 1985

26. Sutherland G, Baker E, Callen D, Campbell H, Young I, Sanderson C, Garson O, Lopez A, Vadas M: Interleukin-5 is at 5q31 and is deleted in the 5q-syndrome. *Blood* 71:1150, 1988

27. Warren D, Moore M: Synergism among interleukin-1, interleukin-3, and interleukin-5 in the production of eosinophils from primitive hemopoietic stem cells. *J Immunol* 140:94, 1988

28. Yang Y-C, Clark S: Molecular cloning of a primate cDNA and the human gene for interleukin-3. *Lymphokines* 15:375, 1988

29. Yang Y-C, Ciarletta A, Temple P, Chung M, Kovacic S, Witte-Grannotti J, Leary A, Kriz R, Donahue R, Wong G, Clark S: Human IL-3 (multi-CSF): Identification by expression cloning of a novel hematopoietic growth factor related to murine IL-3. *Cell* 47:3, 1986

RAPID COMMUNICATION

Activation of the Interleukin-3 Gene by Chromosome Translocation in Acute Lymphocytic Leukemia With Eosinophilia

By Timothy C. Meeker, Dan Hardy, Cheryl Willman, Thomas Hogan, and John Abrams

The t(5;14)(q31;q32) translocation from B-lineage acute lymphocytic leukemia with eosinophilia has been cloned from two leukemia samples. In both cases, this translocation joined the IgH gene and the interleukin-3 (IL-3) gene. In one patient, excess IL-3 mRNA was produced by the leukemic cells. In the second patient, serum IL-3 levels were measured and shown to correlate with disease

activity. There was no evidence of excess granulocyte/macrophage colony stimulating factor (GM-CSF) or IL-5 expression. Our data support the formulation that this subtype of leukemia may arise in part because of a chromosome translocation that activates the IL-3 gene, resulting in autocrine and paracrine growth effects.

© 1990 by The American Society of Hematology.

A NUMBER OF chromosome translocations have been associated with human leukemia and lymphoma. In many cases the study of these translocations has led to the discovery or characterization of proto-oncogenes, such as *bcl-2*, *c-abl*, and *c-myc*, that are located adjacent to the translocation.^{1,2} It is now widely understood that cancer-associated translocations disrupt nearby proto-oncogenes.

A distinct subtype of acute leukemia is characterized by the triad of B-lineage immunophenotype, eosinophilia, and the t(5;14)(q31;q32) translocation.^{3,4} Leukemic cells from such patients have been positive for terminal deoxynucleotidyl transferase (Tdt), common acute lymphoblastic leukemia antigen (CALLA), and CD19, but negative for surface or cytoplasmic immunoglobulin. In previous work, we cloned the t(5;14) breakpoint from one leukemic sample (Case 1) and determined that the IgH and interleukin-3 (IL-3) genes were joined by this abnormality.⁵ In this report, we extend those findings by showing that the t(5;14)(q31;q32) translocation from a second leukemia sample (Case 2) has a similar structure, and we report our study of growth factor expression in these patients.

MATERIALS AND METHODS

Samples and Southern blots. Case 1 has been described.^{5,6} Clinical features of Case 2 have been described in detail.⁷ DNA isolation and Southern blotting was done using previously described methods.⁸ Filters were hybridized with an immunoglobulin Jh probe, a 280 bp *Bam*HI/*Eco*RI genomic IL-3 fragment, and an IL-3 cDNA probe.^{7,8}

Northern blots. RNA isolation and Northern blotting have been described.⁹ Briefly, Northern blots were done by separating 9 µg total RNA on 1% agarose-formaldehyde gels. Equal RNA loading in each lane was confirmed by ethidium bromide staining. Blots were hybridized with an IL-3 cDNA probe extending to the *Xho*I site in exon 5, a 720 bp *Sst*I/*Kpn*I probe derived from intron 2 of the IL-3 gene, a 600 bp *Nhe*I/*Hpa*I IL-5 cDNA probe, and a 500 bp *Pst*I/*Nco*I granulocyte-macrophage colony stimulating factor (GM-CSF) cDNA probe.¹⁰⁻¹²

Polymerase chain reaction. Primers were designed with *Bam*HI sites for cloning. One primer hybridized to the Jh sequences from the IgH gene (Primer 144: 5'-TAGGATCCGACGGTGACAGGGT), and the other hybridized to the region of the TATA box in the IL-3 gene (Primer 161: 5'-AACAGGATCCCGCCTTATATGTGCAG). Polymerase chain reaction (PCR) (95°C for 1 minute, 61°C for 30 seconds, and 72°C for 3 minutes) was done using 500 ng genomic DNA and 50 pmol of each primer in 100 µL containing 67 mmol/L Tris-HCl pH 8.8, 6.7 mmol/L MgCl₂, 10% dimethyl sulfoxide (DMSO), 170 µg/mL bovine serum albumin (BSA) (fraction V),

16.6 mmol/L ammonium sulfate, 1.5 mmol/L each dNTP and Taq polymerase (Perkin-Elmer, Norwalk, CT).¹³

Sequencing. Sequencing was done by chain termination in M13 vectors.¹⁴ As part of this study, we sequenced a subclone of a normal IL-3 promoter, covering 598 base pairs from a *Sma*I site at position -1240 (with respect to the proposed site of transcription initiation) to an *Nhe*I site at position -642. The plasmid containing this region was a gift from Naoko Arai of the DNAX Research Institute.

Expression in Cos7 cells. A genomic IL-3 fragment from Case 1 was cloned into the pXM expression vector.¹⁰ Briefly, the *Hind*III/*Sal*I fragment containing the IL-3 gene was subcloned from the previously described phage clone 4 into pUC18.⁵ The 2.6 kb fragment extending from the *Sma*I site 61 bp upstream of the IL-3 transcription start to the *Sma*I site in the polylinker was cloned into the blunt *Xho*I site of pXM. The negative control construct was the pXM vector without insert. Plasmids were introduced into Cos7 cells by electroporation, and supernatant was collected after 48 hours in culture.

TF1 bioassay. TF-1 cells were passaged in RPMI 1640 supplemented with 10% heat-inactivated fetal bovine serum, 2 mmol L-glutamine, and 1 ng/mL human GM-CSF.¹⁵ Samples and antibodies were diluted in this same medium lacking GM-CSF but containing penicillin and streptomycin. A 25 µL volume of serial dilutions of patient serum was added to wells in a flat bottom 96-well microtiter plate. Rat anti-cytokine monoclonal antibody in a volume of 25 µL was added to appropriate wells and preincubated for 1 hour at 37°C. Fifty microliters of twice washed TF-1 cells were added to each well, giving a final cell concentration of 1 × 10⁵ cells per well (final volume, 100 µL). The plate was incubated for 48 hours. The remaining cell viability was determined metabolically by the colori-

From the Division of Hematology/Oncology 111H, Department of Medicine, University of California and the Veterans Administration Medical Center, San Francisco, CA; the Center for Molecular and Cellular Diagnostics, Department of Pathology and Cell Biology, University of New Mexico, Albuquerque, NM; the Division of Hematology/Oncology, Department of Medicine, West Virginia University, Morgantown, WV; and DNAX Research Institute, Palo Alto, CA.

Submitted March 27, 1990; accepted April 19, 1990.

Supported in part by the University of California Cancer Research Coordinating Committee and University of New Mexico Cancer Center funding from the state of New Mexico. The DNAX Research Institute is supported by Schering-Plough.

Address reprint requests to Timothy C. Meeker, MD, Division of Hematology/Oncology 111H, Department of Medicine, University of California and the Veterans Administration Medical Center, 4150 Clement St, San Francisco, CA 94121.

© 1990 by The American Society of Hematology.

0006-4971/90/7602-0022\$3.00/0

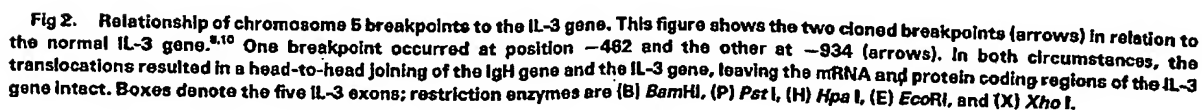
Fig 1. Breakpoint sequences for Case 2. The germline IgJH5 region sequence (protein coding region and recombination signal sequences are underlined) is on top, the translocation sequence from Case 2 (PCR primer sequences and putative N region are underlined) is in the middle, and the germline IL-3 sequence, which we derived from a normal IL-3 clone, is on the bottom.⁷ + indicates that each sequence has the same nucleotide. The sequence documents the head-to-head joining of the IL-3 and IgH genes. The breakpoint in the IL-3 gene occurred at position -834 (*).

Cytokine immunoassays. These assays used rat monoclonal anti-cytokine antibodies (10 µg/mL) to coat the wells of a PVC microtiter plate. The capture antibodies used were: BVD3-6G8, JES1-39D10, and BVD2-23B6, for the IL-3, IL-5, and GM-CSF assays, respectively. Patient sera were then added (undiluted and diluted 1:2 for IL-3, undiluted for IL-5, and undiluted and diluted 1:5 for GM-CSF). The detecting immunoreagents used were either mouse antiserum to IL-3 or nitroiodophenyl (NIP)-derivatized rat monoclonal antibodies JES1-5A2 and BVD2-21C11, specific for IL-5 and GM-CSF, respectively. Bound antibody was subsequently detected with immunoperoxidase conjugates: horseradish peroxidase (HRP)-labeled goat anti-mouse Ig for IL-3, or HRP-labeled rat (J4 MoAb) anti-NIP for IL-5 and GM-CSF. The chromogenic substrate was 3-3'-azino-bis-benzthiazoline sulfonate (ABTS; Sigma, St Louis, MO). Unknown values were interpolated from standard curves prepared from dilutions of the recombinant factors using Softmax software available with the VMAX microplate reader (Molecular Devices).

Leukemic DNA from Case 2 was studied by Southern blotting. When digested with the *Hind*III restriction enzyme and hybridized with a human immunoglobulin heavy chain joining region (Jh) probe, a rearranged fragment at approximately 14 kb was detected (data not shown). When reprobed with either of two different IL-3 probes, a rearranged 14 kb

To characterize better the joining of the IL-3 gene and the immunoglobulin heavy chain (IgH) gene, the polymerase chain reaction (PCR) was used to clone the translocation.¹³ A Jh primer and an IL-3 primer were designed to produce an amplified product in the event of a head-to-head translocation. While control DNA gave no PCR product, Case 2 DNA yielded a PCR-derived fragment of approximately 980 bp, which was cloned and sequenced.

The DNA sequence of the translocation clone from Case 2 confirmed the joining of the Jh region with the promoter of the IL-3 gene in a head-to-head configuration (Fig 1). Sequence analysis indicated that the breakpoint on chromosome 14 was just upstream of the Jh5 coding region. The breakpoint on chromosome 5 occurred 934 bp upstream of the putative site of transcription initiation of the IL-3 gene. We also determined that a putative N sequence of 17 bp was inserted between the chromosome 5 and chromosome 14 sequences during the translocation event.^{17,18} Figure 2 shows



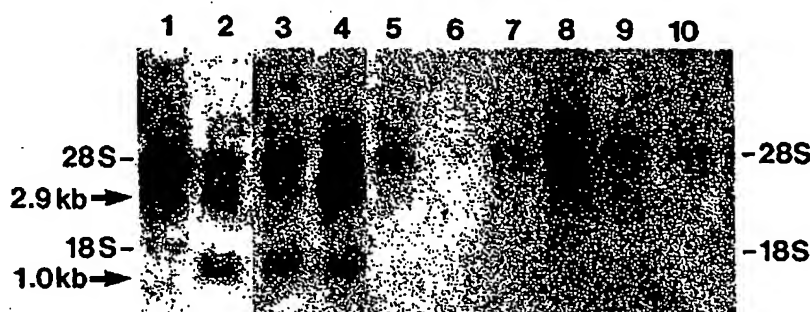


Fig 3. Documentation of IL-3 mRNA over-expression. A Northern blot was prepared and hybridized with a probe for IL-3. Lane 1 contained RNA from unstimulated peripheral blood lymphocytes (PBL) as a negative control. Lane 2 contained RNA from PBL stimulated for 4 hours with concanavalin A (ConA), and lane 3 contained RNA from PBL stimulated with ConA for 48 hours. As in the positive control lanes (2 and 3), a 1 kb band was identified in the leukemic sample from Case 1 (lane 4, lower arrow), suggesting aberrant expression of the IL-3 gene. In addition, the leukemic sample showed over-expression of an unspliced 2.9 kb IL-3 transcript (lane 4, upper arrow). We documented that this represented an unspliced precursor of the mature 1 kb transcript by showing that this band hybridized to a probe from intron 2 of the IL-3 gene. A similar 2.9 kb band was detected in lane 2, suggesting that an IL-3 mRNA of this size is sometimes detectable in normal mitogen-stimulated cells. Lane 5 through 10 represent RNA from six samples of B-lineage acute lymphocytic leukemia without the t(5;14) translocation, indicating that only the sample with the translocation exhibited IL-3 over-expression. Case 2 could not be analyzed by Northern blot because too few cells were available for study.

the locations of the two cloned breakpoints in relation to the IL-3 gene. The two chromosome 5 breakpoints were separated by less than 500 bp.

The genomic structure in Cases 1 and 2 suggested that a normal IL-3 gene product was over-expressed as a result of the altered promoter structure. This would predict that the IL-3 gene on the translocated chromosome was capable of making IL-3 protein. This prediction was tested by expressing a genomic fragment from the translocated allele of Case 1 containing all five IL-3 exons under the control of the SV40 promoter/enhancer in the Cos7 cell line. Cell supernatants were studied in a proliferation assay using the factor dependent erythroleukemic cell line, TF-1. The supernatants derived from transfections using the vector plus insert supported TF-1 proliferation, while supernatants from transfections using the vector alone were negative in this assay (data not shown). Furthermore, the biologic activity could be blocked by an antibody to human IL-3 (BVD3-6G8). This result showed that the translocated allele retained the ability to make IL-3 mRNA and protein.

The level of expression of IL-3 mRNA in leukemic cells from Case 1 was assessed. Northern blotting showed that the mature IL-3 mRNA (approximately 1 kb) and a 2.9 kb unspliced IL-3 mRNA were excessively produced by the leukemia (Fig 3). The 2.9 kb form of the mRNA is also present at low levels in normal peripheral blood T lymphocytes after mitogen activation (Fig 3). Several B-lineage acute leukemia samples without the t(5;14) translocation had undetectable levels of IL-3 mRNA in these experiments. In addition, although genes for GM-CSF and IL-5 map close to the IL-3 gene and might have been deregulated by the translocation, no IL-5 or GM-CSF mRNA could be detected in the leukemic sample (data not shown).^{19,20}

Three serum samples from Case 2 were assayed by immunoassay for levels of IL-3, GM-CSF, and IL-5 (Table 1). Serum IL-3 could be detected and correlated with the clinical course. When the patient's leukemic cell burden was

highest, the IL-3 level was highest. No serum GM-CSF or IL-5 could be detected.

Since the IL-3 immunoassay measured only immunoreactive factor, we confirmed that biologically active IL-3 was present by using the TF-1 bioassay. This bioassay can be rendered monospecific using appropriate neutralizing monoclonal antibodies specific for IL-3, IL-5, or GM-CSF. We observed that sera from 1-16-84 and 3-14-84 contained TF-1 stimulating activity that could be blocked with anti-IL-3 MoAb (BVD3-6G8), but not with MoAbs to IL-5 (JES1-39D10) or GM-CSF (BVD2-23B6) (Fig 4; GM-CSF data not shown). The amount of neutralizable bioactivity in these two samples correlated very well with the difference in IL-3 levels obtained by immunoassay for these samples. Furthermore, the failure to block TF-1 proliferating activity with either anti-IL-5 or anti-GM-CSF was consistent with the inability to measure these factors by immunoassay and

Table 1. Peripheral Blood Counts and Growth Factor Levels at Different Times in Case 2

	Sample Date		
	11/15/83	1/16/84	3/14/84
Peripheral blood counts (cells/ μ L)			
WBC	81,800	116,500	12,300
Lymphoblasts	0	33,785	0
Eosinophils	46,626	73,080	615
Serum growth factor levels (pg/mL)			
IL-3	<444	7,995	1,051
GM-CSF	<15	<15	<15
IL-5	<50	<50	<50

Peripheral blood counts from Case 2 at three different time points with the corresponding growth factor levels quantified by immunoassay. The patient received chemotherapy between 1/16/84 and 3/14/84 to lower his leukemic burden.³ No serum samples were available for a similar analysis of Case 1.

Abbreviation: WBC, white blood cells.

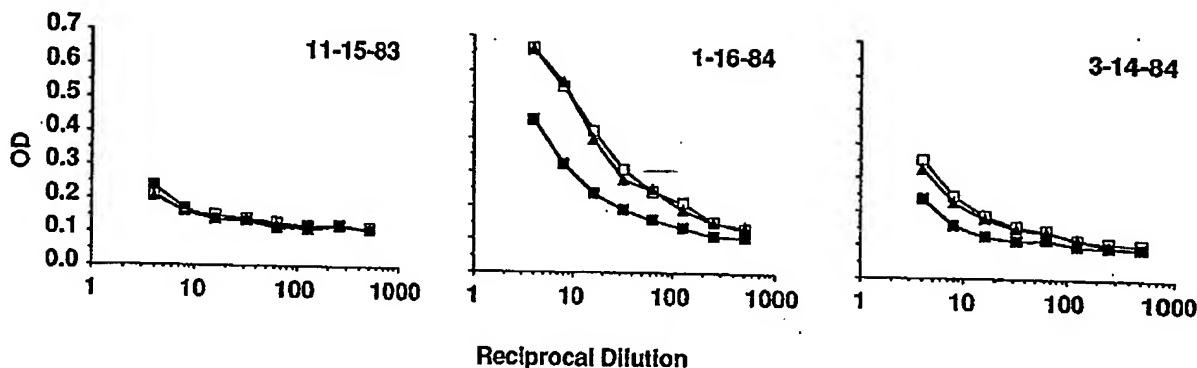


Fig 4. Bioassay of serum IL-3. Leukemic patient sera were tested for bioactive IL-3 and IL-5 in the Tf-1 proliferation assay. The reciprocal of the dilution is indicated on the horizontal axis and the optical density indicating the amount of proliferation is indicated on the vertical axis. Serum from all three time points was assayed simultaneously. The assay was rendered monospecific by using a 1 μ g/mL final concentration of monoclonal rat anti-IL-3, BVD3-6G8 (\blacksquare), or anti-IL-5, JES1-39D10 (\blacktriangle); \square indicates no MoAb. On 1/16/84 and 3/14/84, inhibition of proliferation was evident in the presence of anti-IL-3 antibody, documenting serum levels of IL-3 on those days. Serum IL-5 was not detected in this assay, as anti-IL-5 did not alter Tf-1 proliferation.

indicated that these other myeloid growth factors were not detectably circulating in the serum of this patient.

DISCUSSION

In this report, we have extended our analysis of acute lymphocytic leukemia and eosinophilia associated with the t(5;14) translocation. In both cases we have studied, we have documented the joining of the IL-3 gene from chromosome 5 to the IgH gene from chromosome 14. The breakpoints on chromosome 5 are within 500 bp of each other, suggesting that additional breakpoints will be clustered in a small region of the IL-3 promoter. The PCR assay we have developed will be useful in the screening of additional clinical samples for this abnormality.

The finding of a disrupted IL-3 promoter associated with an otherwise normal IL-3 gene implied that this translocation might lead to the over-expression of a normal IL-3 gene product. In this work, we have documented that this is true. In addition, neither GM-CSF nor IL-5 are over-expressed by the leukemic cells. Furthermore, in one patient, serum IL-3 could be measured and correlated with disease activity. To our knowledge, this is the first measurement of human IL-3 in serum and its association with a disease process. The measurement of serum IL-3 in this and other clinical settings may now be indicated.

The finding of the IL-3 gene adjacent to a cancer-associated translocation breakpoint suggests that its activation is important for oncogenesis. It is our thesis that an autocrine loop for IL-3 is important for the evolution of this leukemia.²¹ The excessive IL-3 production that we have documented would be one feature of such an autocrine loop. The final proof of our thesis must await additional data. In particular, from the study of additional clinical samples, it will be necessary to document that the IL-3 receptor is present on the leukemic cells and that anti-IL-3 antibody decreases proliferation of the leukemia in vitro.

An important aspect of this work is the suggestion of a therapeutic approach for this disease. If an autocrine loop for IL-3 can be documented in this disease, attempts to lower circulating IL-3 levels or block the interaction of IL-3 with its receptor may prove useful. Because it is also possible that the eosinophilia in these patients is mediated by the paracrine effects of leukemia-derived IL-3, similar interventions may improve this aspect of the disease. Antibodies or engineered ligands to accomplish these goals may soon be available.

ACKNOWLEDGMENT

We thank Naoko Arai, Ken-ichi Arai, R. O'Rourke, J. Grimaldi, and T. O'Connell for technical assistance and/or helpful discussions.

REFERENCES

1. Klein G, Klein E: Evolution of tumours and the impact of molecular oncology. *Nature* 315:190, 1985
2. Showe L, Croce C: The role of chromosomal translocations in B- and T-cell neoplasia. *Ann Rev Immunol* 5:253, 1987
3. Hogan T, Koss W, Murgo A, Amato R, Fontana J, VanScoy F: Acute lymphoblastic leukemia with chromosomal 5;14 translocation and hyper eosinophilia: Case report and literature review. *J Clin Oncol* 5:382, 1987
4. Tono-oka T, Sato Y, Matsumoto T, Ueno N, Ohkawa M, Shikano T, Takeda T: Hyper eosinophilic syndrome in acute lymphoblastic leukemia with a chromosome translocation t(5q;14q). *Med Ped Oncol* 12:33, 1984
5. Grimaldi J, Meeker T: The t(5;14) chromosomal translocation in a case of acute lymphocytic leukemia joins the interleukin-3 gene to the immunoglobulin heavy chain gene. *Blood* 73:2081, 1989
6. McConnell T, Foucar K, Hardy W, Saiki J: Three-way reciprocal chromosomal translocation in an acute lymphoblastic leukemia patient with hyper eosinophilia syndrome. *J Clin Oncol* 5:2042, 1987
7. Ravetch J, Siebenlist U, Korsmeyer S, Waldmann T, Leder P: Structure of the human immunoglobulin m locus: Characterization of embryonic and rearranged J and D genes. *Cell* 27:583, 1981
8. Otsuka T, Miyajima A, Brown N, Otsu K, Abrams J, Saeland S, Caux C, Malefijt R, Vries J, Meyerson P, Yokota K, Gemmel L,

- Rennick D, Lee F, Arai N, Arai K, Yokota T: Isolation and characterization of an expressible cDNA encoding human IL-3. *J Immunol* 140:2288, 1988
9. Sambrook J, Fritsch E, Maniatis T: *Molecular Cloning*. Cold Spring Harbor, NY, Cold Spring Harbor Press, 1989
10. Yang Y-C, Ciarletta A, Temple P, Chung M, Kovacic S, Wittek-Giannotti J, Leary A, Kriz R, Donahue R, Wong G, Clark S: Human IL-3 (multi-CSF): Identification by expression cloning of a novel hematopoietic growth factor related to murine IL-3. *Cell* 47:3, 1986
11. Yokota T, Coffman R, Hagiwara H, Rennick D, Takebe Y, Yokota K, Gemmell L, Shrader B, Yang G, Meyerson P, Luh J, Hoy P, Pence J, Briere F, Spits H, Banchereau J, Vries J, Lee F, Arai N, Arai K: Isolation and characterization of lymphokine cDNA clones encoding mouse and human IgA-enhancing factor and eosinophil colony-stimulating factor activities: Relationship to interleukin 5. *Proc Natl Acad Sci USA* 84:7388, 1987
12. Wong G, Wittek J, Temple P, Wilkens K, Leary A, Luxenberg D, Jones S, Brown E, Kay R, Orr E, Shoemaker C, Golde D, Kaufman R, Hewick R, Wang E, Clark S: Human GM-CSF: Molecular cloning of the complementary DNA and purification of the natural and recombinant proteins. *Science* 228:810, 1985
13. Saiki R, Scharf S, Faloona F, Mullis K, Horn G, Erlich H, Arnheim N: Enzymatic amplification of B-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* 230:1350, 1985
14. Norrander U, Kempe T, Messing J: Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 26:101, 1983
15. Kitamura T, Tange T, Terasawa T, Chiba S, Kuwaki T, Miyagawa K, Piao Y, Miyazono K, Urabe A, Takaku F: Establishment and characterization of a unique human cell line that proliferates dependently on GM-CSF, IL-3, or erythropoietin. *J Cell Physiol* 140:323, 1989
16. Mosmann T: Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *J Immunol Methods* 65:55, 1983
17. Bakhshi A, Wright J, Graninger W, Seto M, Owens J, Cossman J, Jensen J, Goldman P, Korsmeyer S: Mechanism of the t(14;18) chromosomal translocation: Structural analysis of both derivative 14 and 18 reciprocal partners. *Proc Natl Acad Sci USA* 84:2396, 1987
18. Tsujimoto Y, Louie E, Bashir M, Croce C: The reciprocal partners of both the t(14;18) and the t(11;14) translocations involved in B-cell neoplasms are rearranged by the same mechanism. *Oncogene* 2:347, 1988
19. Yang Y-C, Kovacic S, Kriz R, Wolf S, Clark S, Welles T, Nienhuis A, Epstein N: The human genes for GM-CSF and IL-3 are closely linked in tandem on chromosome 5. *Blood* 71:958, 1988
20. Sutherland G, Baker E, Callen D, Campbell H, Young I, Sanderson C, Garson O, Lopez A, Vadas M: Interleukin-5 is at 5q31 and is deleted in the 5q- syndrome. *Blood* 71:1150, 1988
21. Sporn M, Roberts A: Autocrine growth factors and cancer. *Nature* 313:745, 1985

Clinical and Pathologic Significance of the *c-erbB-2* (*HER-2/neu*) Oncogene

Timothy P. Singleton and John G. Strickler

The *c-erbB-2* oncogene was first shown to have clinical significance in 1987 by Slamon et al,⁷⁹ who reported that *c-erbB-2* DNA amplification in breast carcinomas correlated with decreased survival in patients with metastasis to axillary lymph nodes. Subsequent studies, however, of *c-erbB-2* activation in breast carcinoma reached conflicting conclusions about its clinical significance. This oncogene also has been reported to have clinical and pathologic implications in other neoplasms. Our review summarizes these various studies and examines the clinical relevance of *c-erbB-2* activation, which has not been emphasized in recent reviews.^{37,38,55} The molecular biology of the *c-erbB-2* oncogene has been extensively reviewed^{37,38,55} and will be discussed only briefly here.

BACKGROUND

The *c-erbB-2* oncogene was discovered in the 1980s by three lines of investigation. The *neu* oncogene was detected as a mutated transforming gene in neuroblastomas induced by ethylnitrosurea treatment of fetal rats.^{8,73,74,75} The *c-erbB-2* was a human gene discovered by its homology to the retroviral gene *v-erbB*.^{33,49,76} *HER-2* was isolated by screening a human genomic DNA library for homology with *v-erbB*.²⁴ When the DNA sequences were determined subsequently, *c-erbB-2*, *HER-2*, and *neu* were found to represent the same gene. Recently, the *c-erbB-2* oncogene also has been referred to as *NGL*.

The *c-erbB-2* DNA is located on human chromosome 17q21^{24,33,68} and codes for *c-erbB-2* mRNA (4.6 kb), which translates *c-erbB-2* protein (p185). This

protein is a normal component of cytoplasmic membranes. The *c-erbB-2* oncogene is homologous with, but not identical to, *c-erbB-1*, which is located on chromosome 7 and codes for the epidermal growth factor receptor.^{2,103} The *c-erbB-2* protein is a receptor on cell membranes and has intracellular tyrosine kinase activity and an extracellular binding domain.^{2,103} Electron microscopy with a polyclonal antibody detects *c-erbB-2* immunoreactivity on cytoplasmic membranes of neoplasms, especially on microvilli and the non-villous outer cell membrane.⁶¹ In normal cells, immunohistochemical reactivity for *c-erbB-2* is frequently present at the basolateral membrane or the cytoplasmic membrane's brush border.^{22,62}

There is experimental evidence that *c-erbB-2* protein may be involved in the pathogenesis of breast neoplasia. Overproduction of otherwise normal *c-erbB-2* protein can transform a cell line into a malignant phenotype.²⁵ Also, when the *neu* oncogene containing an activating point mutation is placed in transgenic mice with a strong promoter for increased expression, the mice develop multiple independent mammary adenocarcinomas.^{18,63} In other experiments, monoclonal antibodies against the *neu* protein inhibit the growth (in nude mice) of a *neu*-transformed cell line,²⁶⁻²⁸ and immunization of mice with *neu* protein protects them from subsequent tumor challenge with the *neu*-transformed cell line.¹⁴ Some authors have speculated that the use of antagonists for the unknown ligand could be useful in future chemotherapy.⁶⁵ Further review of this experimental evidence is beyond the scope of this article.

The *c-erbB-2* activation most likely occurs at an early stage of neoplastic development. This hypothesis is supported by the presence of *c-erbB-2* activation in both in situ and invasive breast carcinomas. In addition, studies of metastatic breast carcinomas usually demonstrate uniform *c-erbB-2* activation at multiple sites in the same patient,^{11,12,39,41,52} although *c-erbB-2* activation has rarely been detected in metastatic lesions but not in the primary tumor.^{57,66,107} Even more rarely, *c-erbB-2* DNA amplification has been detected in a primary breast carcinoma but not in its lymph node metastasis.⁶ In patients who have bilateral breast neoplasms, both lesions have similar patterns of *c-erbB-2* activation, but only a few such cases have been studied.¹¹

MECHANISMS OF *c-erbB-2* ACTIVATION

The most common mechanism of *c-erbB-2* activation is genomic DNA amplification, which almost always results in overproduction of *c-erbB-2* mRNA and protein.^{17,34,65,81} The *c-erbB-2* amplification may stabilize the overproduction of mRNA or protein through unknown mechanisms. Human breast carcinomas with *c-erbB-2* amplification contain 2 to 40 times more *c-erbB-2* DNA^{4,3} and 4 to 128 times more *c-erbB-2* mRNA^{34,90} than found in normal tissue. Most human breast carcinomas with *c-erbB-2* amplification have 2 to 15 times more *c-erbB-2* DNA. Tumors with greater amplification tend to have greater overproduction.^{17,52,65} The non-mammary neoplasms that have been studied tend to have

similar levels of *c-erbB-2* amplification or overproduction relative to the corresponding normal tissue.

The second most common mechanism of *c-erbB-2* activation is overproduction of *c-erbB-2* mRNA and protein without amplification of *c-erbB-2* DNA.⁸¹ The quantities of mRNA and protein usually are less than those in amplified cases and may approach the small quantities present in normal breast or other tissues.^{17,50,52} The *c-erbB-2* protein overproduction without mRNA overproduction or DNA amplification has been described in a few human breast carcinoma cell lines.⁴⁷

Other rare mechanisms of *c-erbB-2* activation have been reported. Translocations involving the *c-erbB-2* gene have been described in a few mammary and gastric carcinomas, although some reported cases may represent restriction fragment length polymorphisms or incomplete restriction enzyme digestions that mimic translocations.^{31,55,75,84,90,103} A single point mutation in the transmembrane portion of *neu* has been described in rat neuroblastomas induced by ethylnitrosourea.^{9,58} The mutated *neu* protein has increased tyrosine kinase activity and aggregates at the cell membrane.^{10,83,98} Although there has been speculation that some of the amplified *c-erbB-2* genes may contain point mutations,⁴⁶ none has been detected in primary human neoplasms.^{41,53,81}

TECHNIQUES FOR DETECTING *c-erbB-2* ACTIVATION

Detection of *c-erbB-2* DNA Amplification

Amplification of *c-erbB-2* DNA is usually detected by DNA dot blot or Southern blot hybridization. In the dot blot method, the extracted DNA is placed directly on a nylon membrane and hybridized with a *c-erbB-2* DNA probe. In the Southern blot method, the extracted DNA is treated with a restriction enzyme, and the fragments are separated by electrophoresis, transferred to a nylon membrane, and hybridized with a *c-erbB-2* DNA probe. In both techniques, *c-erbB-2* amplification is quantified by comparing the intensity (measured by densitometry) of the hybridization bands from the sample with those from control tissue.

Several technical problems may complicate the measurement of *c-erbB-2* DNA amplification. First, the extracted tumor DNA may be excessively degraded or diluted by DNA from stromal cells.⁸¹ Second, the *c-erbB-2* DNA probe must be carefully chosen and labeled. For example, oligonucleotide *c-erbB-2* probes may not be sensitive enough for measuring a low level of *c-erbB-2* amplification, because diploid copy numbers can be difficult to detect (unpublished data). Third, the total amounts of DNA in the sample and control tissue must be compensated for, often with a probe to an unamplified gene. Many studies have used control probes to genes on chromosome 17, the location of *c-erbB-2*, to correct for possible alterations in chromosome number. Identical results, however, are obtained by using control probes to genes on other chromosomes,^{5,55,80} with rare exception.¹⁷ Studies using control probes to the beta-

globin gene must be interpreted with caution, because one allele of this gene is deleted occasionally in breast carcinomas.³

Amplification of *c-erbB-2* DNA was assessed by using the polymerase chain reaction (PCR) in one recent study.³² Oligoprimers for the *c-erbB-2* gene and a control gene are added to the sample's DNA, and PCR is performed. If the sample contains more copies of *c-erbB-2* DNA than of the control gene, the *c-erbB-2* DNA is replicated preferentially.

Detection of *c-erbB-2* mRNA Overproduction

Overproduction of *c-erbB-2* mRNA usually is measured by RNA dot blot or Northern blot hybridization. Both techniques require extraction of RNA but otherwise are analogous to DNA dot blot and Southern blot hybridization. Use of PCR for detection of *c-erbB-2* mRNA has been described in two recent abstracts.^{89,102}

Overproduction of *c-erbB-2* mRNA can be measured by in situ hybridization. Sections are mounted on glass slides, treated with protease, hybridized with a radiolabeled probe, washed, treated with nuclease to remove unbound probe, and developed for autoradiography. Silver grains are seen only over tumor cells that overproduce *c-erbB-2* mRNA. Negative control probes are used.^{85,96,106} Our experience indicates that these techniques are relatively insensitive for detecting *c-erbB-2* mRNA overproduction in routinely processed tissue. Although the sensitivity may be increased by modifications that allow simultaneous detection of *c-erbB-2* DNA and mRNA, in situ hybridization still is cumbersome and expensive (unpublished data).

All of the above *c-erbB-2* mRNA detection techniques have several problems that make them more difficult to perform than techniques for detecting DNA amplification. One major problem is the rapid degradation of RNA in tissue that is not immediately frozen or fixed. In addition, during the detection procedure, RNA can be degraded by RNase, a ubiquitous enzyme, which must be eliminated meticulously from laboratory solutions. Third, control probes to genes that are uniformly expressed in the tissue of interest need to be carefully selected.

Detection of *c-erbB-2* Protein Overproduction

The most accurate methods for detecting *c-erbB-2* protein overproduction are the Western blot method and immunoprecipitation. Both techniques can document the binding specificity of various antibodies against *c-erbB-2* protein. In Western blot studies, protein is extracted from the tissue, separated by electrophoresis (according to size), transferred to a membrane, and detected by using antibodies to *c-erbB-2*. In immunoprecipitation studies, antibodies against *c-erbB-2* are added to a tumor lysate, and the resulting protein-antibody precipitate is separated by gel electrophoresis and stained for protein. Both Western blot and immunoprecipitation are useful research tools but currently are not practical for diagnostic pathology. Two recent abstracts have described an enzyme-linked immunosorbent assay (ELISA) for detection of *c-erbB-2* protein.^{18,45}

Overproduction of c-erbB-2 protein is most commonly assessed by various immunohistochemical techniques. These procedures often generate conflicting results, which are explained at least partially by three factors. First, various studies have used different polyclonal and monoclonal antibodies. Because some polyclonal antibodies recognize weak bands in addition to the c-erbB-2 protein band on Western blot or immunoprecipitation, the results of these studies should be interpreted with caution.^{32,36,47,61} Even some monoclonal antibodies immunoprecipitate protein bands in addition to c-erbB-2 (p185).^{30,59,66} Second, tissue fixation contributes to variability between studies. For example, some antibodies detect c-erbB-2 protein only in frozen tissue and do not react in fixed tissue. In general, formalin fixation diminishes the sensitivity of immunohistochemical methods and decreases the number of reactive cells.^{61,66} When Bouin's fixative is used, there may be a higher percentage of positive cases.⁶² Third, minimal criteria for interpreting immunohistochemical staining are generally lacking. Although there is general agreement that distinct crisp cytoplasmic membrane staining is diagnostic for c-erbB-2 activation in breast carcinoma, the number of positive cells and the staining intensity required to diagnose c-erbB-2 protein overproduction varies from study to study and from antibody to antibody. Degradation of c-erbB-2 protein is not a problem because it can be detected in intact form more than 24 hours after tumor resection without fixation or freezing.⁶⁴

ACTIVATION OF c-erbB-2 IN BREAST LESIONS

Incidence of c-erbB-2 Activation

Most studies of c-erbB-2 oncogene activation do not specify histological subtypes of infiltrating breast carcinoma. Amplification of c-erbB-2 DNA was found in 19.1 percent (519 of 2715) of invasive carcinomas in 25 studies (Table 1), and c-erbB-2 mRNA or protein overproduction was detected in 20.9 percent (566 of 2714) of invasive carcinomas in 20 studies. Twelve studies have documented c-erbB-2 mRNA or protein overproduction in 15 percent (88 of 604) of carcinomas that lacked c-erbB-2 DNA amplification.

The incidence of c-erbB-2 activation in infiltrating breast carcinoma varies with the histological subtype. Approximately 22 percent (142 of 650) of infiltrating ductal carcinomas have c-erbB-2 activation, as expected from the above data. Other variants of breast carcinoma with frequent c-erbB-2 activation are inflammatory carcinoma (62 percent, 54 of 87), Paget's disease (82 percent, 9 of 11), and medullary carcinoma (22 percent, 5 of 23). In contrast, c-erbB-2 activation is infrequent in infiltrating lobular carcinoma (7 percent, 5 of 73) and tubular carcinoma (7 percent, 1 of 15).

The c-erbB-2 protein overproduction is present in 44 percent (44 of 100) of ductal carcinomas in situ and especially comedocarcinoma in situ (68 percent, 49 of 72). The micropapillary type of ductal carcinoma in situ also tends to have c-erbB-2 activation,^{40,54,68} especially if larger cells are present. The greater fre-

TABLE 1. c-erbB-2 ACTIVATION IN MALIGNANT HUMAN BREAST NEOPLASMS

Histological Diagnosis	c-erbB-2 DNA Amplification ^a	c-erbB-2 mRNA Overproduction ^b	c-erbB-2 Protein Overproduction ^c
Carcinoma, not otherwise specified	14/52 ^{a,1} 52/310 ¹⁷	42/180 ^{a,2} 49/126 ^{3d}	118/728 ^{20b}
	52/291 ^{1,10c} 28/176 ⁸⁷	19/62 ^{8d} 19/57 ⁸⁰	58/330 ^{17b} 47/313 ^{8b}
	17/157 ^{11d} 22/141 ²⁵	3/11 ^{1,0c} 6/10 ^{2d} 3/9 ³¹	17/195 ¹¹ 32/191 ^{2b}
	14/136 ⁹⁷ 12/122 ⁴		51/185 ^{10c} 34/102 ^{4c}
	19/103 ^{7a} 15/95 ⁸⁰		24/53 ^{20b} 23/47 ¹⁹
	15/86 ¹¹ⁱ 17/73 ⁷⁷		22/45 ⁸ 11/36 ⁸⁴
	16/66 ^{4c} 6/61 ²⁰		7/24 ⁸¹ 1/10 ⁸¹
	11/57 ^{2c} 10/57 ⁸⁵		
	13/51 ¹³ 8/49 ²¹		
	10/38 ^{2a} 12/36 ⁹⁴		
	1/25 ¹⁵ 7/24 ⁸¹		
	7/15 ³¹ 7/10 ^{2d}		
	2/10 ¹⁰⁷		
Carcinoma, type not specified but lacking c-erbB-2 DNA amplification	—	18/136 ⁸¹ 14/73 ³⁴	16/231 ^{17b} 18/136 ⁸¹
		8/16 ^{8c} 0/8 ⁸⁰ 1/4 ³¹	13/35 ¹³ 14/29 ^{20b}
		0/3 ^{8c}	1/28 ^{2c} 3/24 ⁹⁴
Infiltrating ductal carcinoma			0/17 ⁸¹
	21/118 ^{2c} 23/107 ³⁴	35/85 ³⁴	22/137 ²⁰ 14/93 ⁸⁰
	17/50 ⁴⁴ 7/37 ⁸⁰		9/34 ⁸⁰
Infiltrating ductal carcinoma	14/53 (comedo-carcinoma) ^{11c}		
	3/39 (tubuloductal carcinoma) ¹⁶		

Inflammatory carcinoma	33/80, ³⁵ 3/6 ³²	46/75 ³⁵	5/5 ^{32b}
Paget's disease	—	—	5/5, ⁴⁰ 2/3, ³⁴ 2/2 ³²
Tubular carcinoma	0/5, ¹⁶ 0/1 ³⁰	—	1/9 ⁴⁰
Medullary carcinoma	2/4, ¹⁶ 0/1 ³⁴	0/1 ³⁴	1/12, ⁴⁰ 1/3, ³⁸ 1/2, ³²
Mucinous carcinoma	0/1, ¹⁰ 0/1 ³⁰	—	0/1 ³⁰
Invasive papillary carcinoma	0/2 ³⁰	—	1/2 ³⁰
Infiltrating lobular carcinoma	1/15, ¹⁶ 0/6 ³⁴	1/5 ³⁴	2/27, ³² 0/12, ⁴⁰ 0/9, ³⁰
Mammary fibrosarcoma	0/1 ³⁰	—	1/5 ³⁰
"Benign cystosarcoma"	—	—	—
Ductal CIS ^a with minimal invasion	3/5 ³²	—	0/1 ³⁰
Ductal CIS	0/2 ³⁴	1/2 ³⁴	—
Ductal CIS, solid or comedo type	—	—	33/74, ⁴⁰ 10/24, ³⁰
Ductal CIS, micropapillary type	—	—	20/33, ³⁴ 19/29, ³²
Ductal CIS, micropapillary or cribriform type	—	—	10/10 ³⁴
Ductal CIS, papillary or cribriform type	—	—	10/10 ³⁰
Lobular CIS	—	—	1/(local)/14 ³⁴
	—	—	0/16, ³² 1/9, ³⁰ 0/3 ⁴⁰
	—	—	0/16 ⁴⁰

^aShown as number of cases with activation/number of cases studied; reference is given as a superscript.

^bThese protein studies used Western blots; the rest used immunohistochemical methods.

^cCIS = carcinoma in situ.

quency of *c-erbB-2* protein overproduction in comedocarcinoma in situ, compared with infiltrating ductal carcinoma, could be explained by the fact that many infiltrating ductal carcinomas arise from other types of intraductal carcinoma, which show *c-erbB-2* activation infrequently. Others have speculated that carcinoma in situ with *c-erbB-2* activation tends to regress or to lose *c-erbB-2* activation during progression to invasion.^{40,63,82} Infiltrating and in situ components of ductal carcinoma, however, usually are similar with respect to *c-erbB-2* activation,^{11,39} although some authors have noted more heterogeneity of the immunohistochemical staining pattern in invasive than in in situ carcinoma.^{40,42,63} Activation of *c-erbB-2* is infrequent in lobular carcinoma in situ. If lesions contain more than one histological pattern of carcinoma in situ, the *c-erbB-2* protein overproduction tends to occur in the comedocarcinoma in situ but may include other areas of carcinoma in situ.^{42,54,63} Overproduction of *c-erbB-2* protein in ductal carcinoma in situ correlates with larger cell size and a periductal lymphoid infiltrate.⁶³

Activation of *c-erbB-2* has not been identified in benign breast lesions, including fibrocystic disease, fibroadenomas, and radial scars (Table 2). Strong membrane immunohistochemical reactivity for *c-erbB-2* has not been described in atypical ductal hyperplasia, although weak accentuation of membrane staining has been noted infrequently.^{39,42,54} In normal breast tissue, *c-erbB-2* DNA is diploid, and *c-erbB-2* is expressed at lower levels than in activated tumors.^{34,35,65,83}

These preliminary data suggest that *c-erbB-2* activation may not be useful for resolving many of the common problems in diagnostic surgical pathology. For example, *c-erbB-2* activation is infrequent in tubular carcinoma and radial scars. In addition, because *c-erbB-2* activation is unusual in atypical ductal hyperplasia, cribriform carcinoma in situ, and papillary carcinoma in situ, detection of *c-erbB-2* activation in these lesions may not be helpful in their differential diagnosis. The histological features of comedocarcinoma in situ, which commonly overproduces *c-erbB-2*, are unlikely to be mistaken for those of benign lesions. Activation of

TABLE 2. *c-erbB-2* ACTIVATION IN BENIGN HUMAN BREAST LESIONS

Histological Diagnosis	<i>c-erbB-2</i> DNA Amplification ^a	<i>c-erbB-2</i> mRNA Overproduction	<i>c-erbB-2</i> Protein Overproduction
Fibrocystic disease	0/10 ³³	—	0/32, ³⁹ 0/9, ⁶³ 0/8 ⁶⁵
Atypical ductal hyperplasia	—	—	2(weak)/21, ⁵⁴ 1(cytoplasmic)/13 ³⁹
Benign ductal hyperplasia	—	—	0/12 ³⁹
Sclerosing adenosis	—	—	0/4 ³⁹
Fibroadenomas	0/16, ³⁴ 0/6, ⁶³ 0/2, ⁵¹ 0/1 ⁹¹	0/6, ³⁵ 0/3 ³⁴	0/21, ³⁹ 0/10, ⁶³ 0/8, ⁶⁵ 0/3 ⁴²
Radial scars	—	—	0/22 ³⁹
Blunt duct adenosis	—	—	0/14 ³⁹
"Breast mastosis"	—	0/3 ³⁵	—

^aShown as number of cases with activation/number of cases studied; reference is given as a superscript.

c-erbB-2, however, does favor infiltrating ductal carcinoma over infiltrating lobular carcinoma. Further studies of these issues would be useful.

Correlation of c-erbB-2 Activation With Pathologic Prognostic Factors

Multiple studies have attempted to correlate c-erbB-2 activation with various pathologic prognostic factors (Table 3). Activation of c-erbB-2 was correlated with lymph node metastasis in 8 of 28 series, with higher histological grade in 6 of 17 series, and with higher stage in 4 of 14 series. Large tumor size was not associated with c-erbB-2 activation in most studies (11 of 14). Tetraploid DNA content and low proliferation, measured by Ki-67, have been suggested as prognostic factors and may correlate with c-erbB-2 activation.^{6,7}

Correlation of c-erbB-2 Activation With Clinical Prognostic Factors

Various studies have attempted also to correlate c-erbB-2 activation with clinical features that may predict a poor outcome (Table 4). Activation of c-erbB-2 correlated with absence of estrogen receptors in 10 of 28 series and with absence of progesterone receptors in 6 of 18 series. In most studies, patient age did not correlate with c-erbB-2 activation, and, in the rest of the reports, c-erbB-2 activation was associated with either younger or older ages.

Correlation of c-erbB-2 Activation With Patient Outcome

Slamon et al^{79,81} first showed that amplification of the c-erbB-2 oncogene independently predicts decreased survival of patients with breast carcinoma. The correlation of c-erbB-2 amplification with poor outcome was nearly as strong as the correlation of number of involved lymph nodes with poor outcome. Slamon et al also reported that c-erbB-2 amplification is an important prognostic indicator only in patients with lymph node metastasis.^{79,81}

A large number of subsequent studies also attempted to correlate c-erbB-2 activation with prognosis (Table 5). In 12 series, there was a correlation between c-erbB-2 activation and tumor recurrence or decreased survival. In five of these series, the predictive value of c-erbB-2 activation was reported to be independent of other prognostic factors. In contrast, 18 series did not confirm the correlation of c-erbB-2 activation with recurrence or survival. Four possible explanations for this controversy are discussed below.

One problem is that c-erbB-2 amplification correlates with prognosis mainly in patients with lymph node metastasis. As summarized in Table 5, most studies of patients with axillary lymph node metastasis showed a correlation of c-erbB-2 activation with poor outcome. In contrast, most studies of patients without axillary metastasis have not demonstrated a correlation with patient outcome. Table 6 summarizes the studies in which all patients (with and without axillary metastasis) were considered as one group. There is a trend for studies with a higher percentage of metastatic cases to show an association between c-erbB-2 activation and poor outcome. Thus, most of the current evidence suggests that c-erbB-2 activation has prognostic value only in patients with metastasis to lymph nodes.

TABLE 3. CORRELATION OF c-erbB-2 ACTIVATION WITH PATHOLOGIC PROGNOSTIC FACTORS IN BREAST CARCINOMA

Prognostic Factor	P ^a	c-erbB-2 DNA Amplification ^b	c-erbB-2 mRNA Overproduction	c-erbB-2 Protein Overproduction ^c
Metastasis to axillary lymph nodes	<0.05 0.05-0.15 >0.15	(118) ³⁵ (105) ³⁴ (49) ²¹ (103) ³⁵ (86) ³² (58) ¹¹¹ (279) ¹⁷ (176) ⁸⁷ (157) ¹¹³ (122) ⁴ (85) ³⁰ (50) ³² (50) ⁴⁴ (47) ¹³ (41) ³³	(104) ³⁵ (92) ³⁴ (9) ²¹ — (50) ³⁰	(350) ³⁵ (36) ¹³ (189) ³² (329) ¹⁷⁰ (261) ³⁵ (195) ¹¹ (185) ¹⁰¹ (102) ³⁹ (50) ³²⁰
Larger size	<0.05 0.05-0.15 >0.15	(280) ¹⁷ (96) ⁷⁶ (176) ⁸⁷ (157) ¹¹³ (103) ¹⁷⁰ (64) ⁷⁷ (58) ¹¹¹ (45) ²¹	— — (51) ³⁰	(330) ¹⁷⁰ (189) ³² — (350) ³⁵ (185) ¹⁰¹ (34) ³²
Higher stage	<0.05 0.05-0.15 >0.15	(300) ¹⁷ (64) ⁷⁷ (58) ¹¹¹ (56) ³² (176) ⁸⁷ (157) ¹¹³ (84) ³⁰ (61) ³⁰ (53) ²¹ (52) ⁸⁷ (41) ³⁰	— — — —	(349) ¹⁷⁰ — (102) ³⁹ (56) ³²⁰
Higher histological grade	<0.05 0.05-0.15 >0.15	(47) ¹³ (15) ³¹ — (122) ⁴ (113) ³⁴ (95) ³⁰ (58) ¹¹¹ (50) ⁴⁴ (41) ³³	(53) ³⁵ — (86) ³⁵ (65) ³⁵	(176) ¹⁰¹ (169) ¹¹ (38) ¹³ — (290) ³⁵ (189) ³² (102) ³⁹

^aA correlation is statistically significant at <0.05, equivocal at best between 0.05 and 0.15, and not statistically significant at >0.15.^bNumbers inside parentheses are the number of patients in an individual study; superscript is the reference. Some studies analyzed more than one group of patients.^cby Western blot method; all other protein studies used immunohistochemical methods.

TABLE 4. CORRELATION OF c-erbB-2 ACTIVATION WITH CLINICAL PROGNOSTIC FACTORS IN BREAST CARCINOMA

Prognostic Factor	P ^a	c-erbB-2 DNA Amplification ^b	c-erbB-2 mRNA Overproduction ^c	c-erbB-2 Protein Overproduction ^c
Absence of estrogen receptors	<0.05	(253) ¹⁰⁹ (141) ³⁵ (109) ³⁴ (86) ⁷⁸ (50) ⁴⁴ (47) ¹³	(104) ³⁵	(350) ^{35c} (330) ^{17c} (185) ¹⁰¹
	0.05-0.15	(157) ¹¹³ (122) ⁴ (103) ⁷⁹ (95) ³⁰ (64) ⁷⁷ (61) ⁸⁰	(180) ³⁵ (82) ³⁵ (62) ³⁵ (57) ³⁰	(290) ^{35c} (172) ¹¹ (51) ^{52c} (38) ¹³
	>0.15	(58) ¹¹¹ (53) ²¹ (51) ⁸² (41) ³²		
Absence of progesterone receptors	<0.05	(253) ¹⁰⁹ (141) ³⁵ (109) ³⁴ (50) ⁴⁴	—	(350) ^{35c} (306) ^{17c}
	0.05-0.15	(86) ⁷⁸ (49) ⁸² (157) ¹¹³ (122) ⁴ (103) ⁷⁹	(180) ³⁵ (103) ³⁵ (82) ³⁵ (56) ³⁵	(90) ¹¹ (49) ^{52c}
	>0.15	(64) ⁷⁷	—	—
Age (menopausal status)	<0.05	—	—	(younger: 330) ^{17c} (older: 56) ^{52c}
	0.05-0.15	(younger: 86) ⁷⁸ (230) ¹⁷ (178) ⁸⁷ (157) ¹¹³	(62) ³⁵	(350) ^{35c} (280) ³⁵ (189) ³² (162) ¹¹ (45) ⁸²
	>0.15	(122) ⁴ (116) ³⁴ (103) ⁷⁹ (95) ³⁰ (64) ⁷⁷ (58) ¹¹ (56) ³² (53) ²¹ (49) ¹³ (41) ³² (15) ³¹	—	—

^aA correlation is statistically significant at <0.05, equivocal at best between 0.05 and 0.15, and not statistically significant at >0.15^bNumbers inside parentheses are the number of patients in an individual study; superscript is the reference. Some studies analyzed more than one group of patients.^cBy Western blot method; all other protein studies used immunohistochemical methods.

TABLE 5. CORRELATION OF c-erbB-2 ACTIVATION WITH OUTCOME IN PATIENTS WITH BREAST CARCINOMA

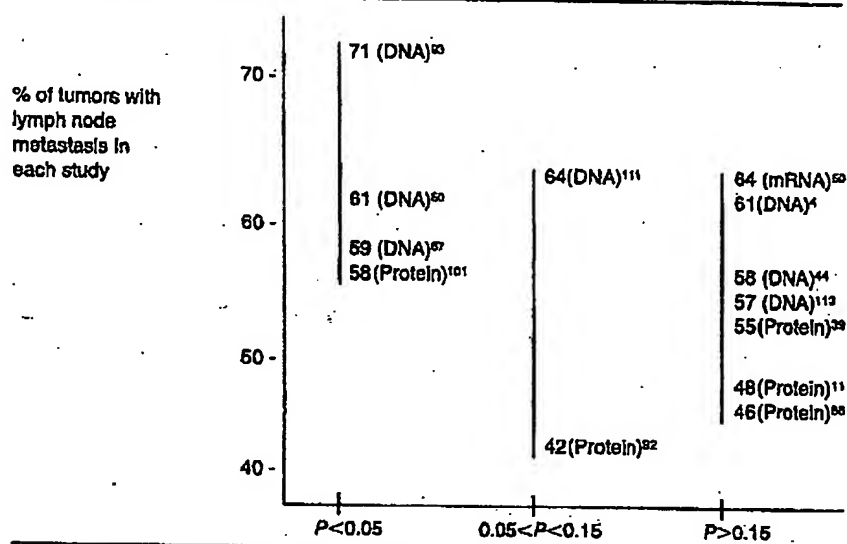
P ^a	Type of c-erbB-2 Activation ^b	Number of Patients		Statistical Analysis ^c	Reference
		Total	With Metastasis to Axillary Lymph Nodes		
<0.05	DNA	176		M	87
<0.05	DNA	61		U	60
<0.05	DNA	57		U	65
<0.05	DNA	41		U	93
<0.05	mRNA	62		U	65
<0.05	Protein	102		M	101
<0.05	DNA		345	M	81
<0.05	DNA		120	U	17
<0.05	DNA		91	U	87
<0.05	DNA		86	M	79
<0.05	Protein-WB		350	M	85
<0.05	Protein		62	U	101
0.05-0.15	DNA	57		U	111
0.05-0.15	Protein	189		M	92
0.05-0.15	Protein		120	U	86
>0.15	DNA	130		U	113
>0.15	DNA	122		M	4
>0.15	DNA	50		U	44
>0.15	mRNA	57		U	50
>0.15	Protein	290		M	88
>0.15	Protein	195		U	11
>0.15	Protein	102		U	39
>0.15	Protein		137	U	17
>0.15	DNA			M	81
>0.15	DNA			U	17
>0.15	DNA			U	87
>0.15	Protein-WB			U	85
>0.15	Protein-WB			U	17
>0.15	Protein			U	86
>0.15	Protein			U	40

^aThe endpoints of these studies were tumor recurrence or decreased survival or both. Correlation between c-erbB-2 activation and a poorer patient outcome is statistically significant at <0.05, is of equivocal significance at 0.05 to 0.15, and is not significant at >0.15.

^bShown as variable measured. Letters "WB" indicate assay by Western blot; the other protein studies used immunohistochemical methods.

^cM = multivariate statistical analysis; U = univariate statistical analysis.

TABLE 6. PERCENTAGE OF BREAST CARCINOMAS WITH METASTASIS COMPARED WITH PROGNOSTIC SIGNIFICANCE OF c-erbB-2 ACTIVATION



P for correlation of c-erbB-2 activation with patient outcome.

Each study's percentage of breast carcinomas with metastasis is compared with the correlation between c-erbB-2 activation and outcome. These data include only those studies that considered, as one group, all breast cancer patients, whether or not they had axillary metastasis. Superscripts are the references. In parentheses are the types of c-erbB-2 activation. P values are interpreted as in Table 3.

A second problem is that various types of breast carcinoma are grouped together in many survival studies. Because the current literature suggests that c-erbB-2 activation is infrequent in lobular carcinoma, studies that combine infiltrating ductal and lobular carcinomas may dilute the prognostic effect of c-erbB-2 activation in ductal tumors. In addition, most studies do not analyze inflammatory breast carcinoma separately. This condition frequently shows c-erbB-2 activation and has a worse prognosis than the usual mammary carcinoma, but it is an uncommon lesion.

A third potential problem is the paucity of studies that attempt to correlate c-erbB-2 activation with clinical outcome in subsets of breast carcinoma without metastasis. Two recent abstracts reported that in patients without lymph node metastasis who had various risk factors for recurrence (such as large tumor size and absence of estrogen receptors), c-erbB-2 overexpression predicted early recurrence.^{23,57} In patients with ductal carcinoma in situ, one small study found no association between tumor recurrence and c-erbB-2 activation.⁴⁰

A fourth problem is the lack of data regarding whether the prognosis correlates better with c-erbB-2 DNA amplification or with mRNA or protein overproduction. Most studies that find a correlation between c-erbB-2 activa-

tion and poor patient outcome measure *c-erbB-2* DNA amplification (Table 5), and breast carcinoma patients with greater amplification of *c-erbB-2* may have poorer survival.^{79,81} Recent studies suggest that amplification has more prognostic power than overproduction,^{17,34,35} but the clinical significance of *c-erbB-2* overproduction without DNA amplification deserves further research.^{17,53} Few studies have attempted to correlate patient outcome with *c-erbB-2* mRNA overproduction, and many studies of *c-erbB-2* protein overproduction use relatively less reliable methods such as immunohistochemical studies with polyclonal antibodies.

Comparison of *c-erbB-2* Activation With Other Oncogenes in Breast Carcinoma

Other oncogenes that may have prognostic implications in human breast cancer are reviewed elsewhere.^{71,106} This section will be restricted to a comparison between the clinical relevance of *c-erbB-2* and these other oncogenes.

The *c-myc* gene is often activated in breast carcinomas, but *c-myc* activation generally has less prognostic importance than *c-erbB-2* activation.^{21,34,77,87,93} One study found a correlation between increased mRNAs of *c-erbB-2* and *c-myc*, although other reports have not confirmed this.^{34,106} Subsequent research, however, could demonstrate a subset of breast carcinomas in which *c-myc* has more prognostic importance than *c-erbB-2*.

The gene *c-erbB-1* for the epidermal growth factor receptor (EGFR) is homologous with *c-erbB-2* but is infrequently amplified in breast carcinomas.⁷⁹ Overproduction of EGFR, however, occurs more frequently than amplification and may correlate with a poor prognosis. In studies that have examined both *c-erbB-2* and EGFR in the same tumor, *c-erbB-2* has a stronger correlation with poor prognostic factors.^{35,52} Studies have tended to show no correlation between amplification of *c-erbB-2* and *c-erbB-1* or overproduction of *c-erbB-2* and EGFR, although at the molecular level EGFR mediates phosphorylation of *c-erbB-2* protein.^{51,52,61,69,100} Recent reviews describe EGFR in breast carcinoma.^{43,100}

The genes *c-erbA* and *ear-1* are homologous to the thyroid hormone receptor, and they are located adjacent to *c-erbB-2* on chromosome 17. These genes are frequently coamplified with *c-erbB-2* in breast carcinomas. The absence of *c-erbA* expression in breast carcinomas, however, is evidence against an important role for this gene in breast neoplasia.⁹⁰ Amplification of *c-erbB-2* can occur without *ear-1* amplification, and these tumors have a decreased survival that is similar to tumors with both *c-erbB-2* and *ear-1* amplification.⁹⁷ Consequently, *c-erbB-2* amplification seems to be more important than amplification of *c-erbA* or *ear-1*.

Other genes also have been compared with *c-erbB-2* activation in breast carcinomas. One study found a significant correlation between increased *c-erbB-2* mRNA and increased mRNAs of *fos*, platelet-derived growth factor chain A, and *Ki-ras*.¹⁰⁶ Allelic deletion of *c-Ha-ras* may indicate a poorer prognosis in breast carcinoma,²¹ but it has not been compared with *c-erbB-2* activation. Some studies have suggested a correlation between advanced stage or recurrence of breast carcinoma and activation of any one of several oncogenes.^{21,113}

ACTIVATION OF c-erbB-2 IN NON-MAMMARY TISSUES

Incidence of c-erbB-2 Activation in Non-Mammary Tissues

Table 7 summarizes the normal tissues in which c-erbB-2 expression has been detected, usually with immunohistochemical methods using polyclonal anti-

TABLE 7. PRESENCE OR ABSENCE OF c-erbB-2 mRNA OR c-erbB-2 PROTEIN IN NORMAL HUMAN TISSUES

Tissues With c-erbB-2 mRNA	Tissues Producing c-erbB-2 Protein ^a	Tissues Lacking c-erbB-2 mRNA	Tissues Lacking c-erbB-2 Protein
Skin ²⁴	Epidermis ⁵⁰ External root sheath ⁵⁰ Eccrine sweat gland ⁵⁰ Fetal oral mucosa ⁶² Fetal esophagus ⁶²		Postnatal oral mucosa ⁶² Postnatal esophagus ⁶²
Stomach ²⁴	Stomach ^{22,62} Fetal intestine ^{52a}		
Jejunum ²⁴	Small intestine ^{22,62}		
Colon ²⁴	Colon ^{22,62}		
Kidney ²⁴	Fetal kidney ^{62a}	Kidneys ¹⁰⁴	Glomerulus ⁶² Postnatal Bowman's capsule ⁶² Postnatal proximal tubule ⁶² Postnatal collecting duct ⁶² Postnatal renal pelvis ⁶² Postnatal fetal ureter ⁶²
	Fetal proximal tubule ⁶² Distal tubule ⁶² Fetal collecting duct ⁶² Fetal renal pelvis ⁶² Fetal ureter ⁶²		Liver ^{62,65}
Liver ²⁴	Hepatocytes ²² Pancreatic acini ²² Pancreatic ducts ^{22,62} Endocrine cells of islets of Langerhans ²²		Pancreatic islets ⁶²
Lung ²⁴	Fetal trachea ⁶² Fetal bronchioles ⁶² Bronchioles ⁵⁰		Postnatal trachea ⁶² Postnatal bronchioles ⁶² Postnatal alveoli ^{62,59} Postnatal brain ⁶² Postnatal ganglion cells ⁶²
Fetal brain ²⁴	Fetal ganglion cells ⁶²		
Thyroid ¹			
Uterus ²⁴	Ovary ¹² Blood vessels ⁴²		Endothelium ⁶²
Placenta ²⁴			Adrenocortical cells ⁶² Postnatal thymus ⁶² Fibroblasts ⁶² Smooth muscle cells ⁶² Cardiac muscle cells ⁶²

^aThis protein study used Western blots; the rest used immunohistochemical methods.

bodies. Only a few studies have been performed, and some of these do not demonstrate convincing cell membrane reactivity in the published photographs. The interpretations in these studies, however, are listed, with the caveat that these findings should be confirmed by immunoprecipitation or Western or RNA blots. Production of *c-erbB-2* has been identified in normal epithelium of the gastrointestinal tract and skin. Discrepancies regarding *c-erbB-2* protein in other tissues could be due, at least in part, to differences in techniques.

The data on *c-erbB-2* activation in various non-mammary neoplasms should be interpreted with caution, because only small numbers of tumors have been studied, usually by immunohistochemical methods using polyclonal antibodies. Studies using cell lines have been excluded, because cell culture can induce amplification and overexpression of other genes, although this has not been documented for *c-erbB-2*.

Activation of *c-erbB-2* has been identified in 32 percent (64 of 203) of ovarian carcinomas in eight studies (Table 8). One abstract⁴⁵ stated that ovarian carcinomas contained significantly more *c-erbB-2* protein than ovarian non-epithelial malignancies. Another report⁶¹ showed that 12 percent of ovarian carcinomas had *c-erbB-2* overproduction without amplification.

Activation of *c-erbB-2* has been identified in 20 percent (40 of 188) of gastric adenocarcinomas in seven studies, including 33 percent (21 of 64) of

TABLE 8. *c-erbB-2* ACTIVATION IN HUMAN GYNECOLOGIC TUMORS*

Tumor Type	<i>c-erbB-2</i> DNA Amplification	<i>c-erbB-2</i> mRNA Over-production	<i>c-erbB-2</i> Protein Over-production
Ovary—carcinoma, not otherwise specified	31/120, ⁵¹ 1/11, ⁵⁷ 0/5, ¹⁰⁷ 0/5, ⁸⁴ 0/3, ¹¹² 0/2, ⁷³ 0/1 ¹¹⁰	23/67 ⁶¹	23/73, ¹² 36/72 ⁶¹
Ovary—serous (papillary) carcinoma	2/7, ¹¹⁰ 1/7, ¹¹² 0/5 ⁷²	—	—
Ovary—endometrioid carcinoma	0/3 ¹¹⁰	—	—
Ovary—mucinous carcinoma	1/2, ¹¹⁰ 0/1 ⁷²	—	—
Ovary—clear cell carcinoma	0/2, ¹¹² 0/1 ⁷²	—	—
Ovary—mixed epithelial carcinoma	0/2 ⁷²	—	—
Ovary—endometrioid borderline tumor	0/1 ⁷²	—	—
Ovary—mucinous borderline tumor	0/3 ⁷²	—	—
Ovary—serous cystadenoma	0/4 ⁷²	—	—
Ovary—mucinous cystadenoma	0/2 ⁷²	—	—
Ovary—sclerosing stromal tumor	0/1 ⁷²	—	—
Ovary—fibrothecoma	0/1 ⁷²	—	—
Uterus—endometrial adenocarcinoma	0/4, ⁸⁴ 0/1 ¹¹⁰	—	—

*Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. All protein studies used immunohistochemical methods.

intestinal or tubular subtypes and 9 percent (4 of 47) of diffuse or signet ring cell subtypes (Table 9). Activation of c-erbB-2 has been detected in 2 percent (6 of 281) of colorectal carcinomas, although an additional immunohistochemical study detected c-erbB-2 protein in seven of eight tissues fixed in Bouin's solution. One study found greater immunohistochemical reactivity for c-erbB-2 protein in colonic adenomatous polyps than in the adjacent normal epithelium, using Bouin's fixative. Lesions with anaplastic features and progression to invasive carcinoma tended to show decreased immunohistochemical reactivity for c-erbB-2 protein.²² Hepatocellular carcinomas (12 of 14 cases) and cholangiocarcinomas (46 of 63 cases) reacted with antibodies against c-erbB-2 in one study, but some of these "positive" cases showed only diffuse cytoplasmic staining, which

TABLE 9. c-erbB-2 ACTIVATION IN HUMAN GASTROINTESTINAL TUMORS*

Tumor Type	c-erbB-2 DNA Amplification	c-erbB-2 Protein Overproduction
Esophagus—squamous cell carcinoma	0/1 ¹⁰⁷	0/1 ⁸¹
Stomach—carcinoma, poorly differentiated	0/22 ¹⁰⁸	—
Stomach—adenocarcinoma	2/24, ⁸⁴ 2/9, ¹⁰⁷ 2/8, ¹¹¹ 2/8, ⁸⁷ 0/1 ¹⁰⁸	4/27, ²² 3/10 ⁸¹
Stomach—carcinoma, intestinal or tubular type	5/10 ¹⁰⁸	16/54 ²²
Stomach—carcinoma, diffuse or signet ring cell type	0/2 ¹⁰⁸	4/45 ²²
Colorectum—carcinoma	2/49, ⁸⁴ 1/45, ¹¹¹ 1/45, ⁸⁷ 1/45, ⁸⁰ 0/40, ⁸¹ 0/32, ¹⁰⁷ 0/3 ⁸²	1/22, ⁵⁸ 7/8 ^{22b}
Colon—villous adenoma	0/1 ⁸⁰	—
Colon—tubulovillous adenoma	0/5 ⁸⁰	—
Colon—tubular adenoma	0/7 ⁸⁰	19/19 ^{22b}
Colon—hyperplastic polyp	0/1 ⁸⁰	—
Intestine—leiomyosarcoma	—	0/1 ⁸¹
Hepatocellular carcinoma	0/12 ¹¹¹	12/14, ⁸⁵ 0/2 ⁸¹
Hepatoblastoma	0/1 ⁸⁷	—
Cholangiocarcinoma	—	46/63 ²²
Pancreas—adenocarcinoma	—	2/80, ^{41c} 0/2 ⁸¹
Pancreas—acinar carcinoma	—	0/1 ⁴¹
Pancreas—clear cell carcinoma	—	0/2 ⁴¹
Pancreas—large cell carcinoma	—	0/3 ⁴¹
Pancreas—signet ring carcinoma	—	0/1 ⁴¹
Pancreas—chronic inflammation	—	0/14 ^{41c}

*Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. All protein studies used immunohistochemical methods. No studies analyzed for c-erbB-2 mRNA.

^bTissues fixed in Bouin's solution.

^cOnly cases with distinct membrane staining are interpreted as showing c-erbB-2 overproduction.

TABLE 10. *c-erbB-2* ACTIVATION IN HUMAN PULMONARY TUMORS*

Tumor Type	<i>c-erbB-2</i> DNA Amplification	<i>c-erbB-2</i> Protein Overproduction
Non-small cell carcinoma	2/60, ⁷⁵ 0/60 ⁸¹	1/84 ⁸³
Epidermoid carcinoma	0/13, ⁸² 0/10, ⁸⁷ 0/6 ⁸⁰	3/5 ⁸⁰
Adenocarcinoma	0/21, ⁸² 1/13, ⁸⁰ 0/7, ¹¹¹ 0/7, ⁸⁷ 0/3 ¹⁰⁷	4/12 ⁸⁰
Large cell carcinoma	0/9, ⁸² 0/6 ⁸⁰	—
Small cell carcinoma	—	0/26, ⁸⁸ 0/3 ⁸⁰
Carcinoid tumor	0/1 ⁸²	0/3 ⁸⁰

*Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. All protein studies used immunohistochemical methods. No studies analyzed for *c-erbB-2* mRNA.

does not indicate *c-erbB-2* activation in breast neoplasms.⁸⁵ Also, some pancreatic carcinomas and chronic pancreatitis tissue had cytoplasmic immunohistochemical reactivity for *c-erbB-2* protein, in addition to the rare case of pancreatic adenocarcinoma with distinct cell membrane staining.⁴¹

Tables 10 through 14 summarize the studies of *c-erbB-2* activation in other neoplasms. The *c-erbB-2* oncogene is not activated in most of these tumors. Activation of *c-erbB-2* has been detected in 1 percent (4 of 299) of pulmonary non-small cell carcinomas in nine studies, although one additional report⁸⁹ found *c-erbB-2* protein overproduction in 41 percent (7 of 17). Renal cell carcinoma had *c-erbB-2* activation in 7 percent (2 of 30) in four studies. Overproduction of *c-erbB-2* protein was described in one transitional cell carcinoma of the urinary bladder, a grade 2 papillary lesion.⁸⁸ Squamous cell carcinoma and basal cell carcinoma of the skin may contain *c-erbB-2* protein, but it is not clear

TABLE 11. *c-erbB-2* ACTIVATION IN HUMAN HEMATOLOGIC PROLIFERATIONS*

Tumor Type	<i>c-erbB-2</i> DNA Amplification	<i>c-erbB-2</i> mRNA Overproduction	<i>c-erbB-2</i> Protein Overproduction
Hematologic malignancies	0/23 ¹¹¹	—	—
Malignant lymphoma	0/9, ⁸⁷ 0/3 ¹⁰⁷	0/1 ¹	0/15 ⁸¹
Acute leukemia	0/14 ⁸⁷	—	—
Acute lymphoblastic leukemia	0/1 ¹⁰⁷	—	—
Acute myeloblastic leukemia	0/3 ¹⁰⁷	—	—
Chronic leukemia	0/19 ⁸⁷	—	—
Chronic lymphocytic leukemia	0/6 ¹⁰⁷	—	—
Chronic myelogenous leukemia	0/8 ¹⁰⁷	—	—
Myeloproliferative disorder	0/1 ⁸⁷	—	—

*Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. All protein studies used immunohistochemical methods.

TABLE 12. c-erbB-2 ACTIVATION IN HUMAN TUMORS OF SOFT TISSUE AND BONE*

Tumor Type	c-erbB-2 DNA Amplification
Sarcoma	0/10, ¹¹¹ 0/8 ⁸⁷
Malignant fibrous histiocytoma	0/1 ¹⁰⁷
Liposarcoma	0/3 ¹⁰⁷
Pleomorphic sarcoma	0/1 ¹⁰⁷
Rhabdomyosarcoma	0/1 ¹⁰⁷
Osteogenic sarcoma	0/2, ¹⁰⁷ 0/2 ⁸⁷
Chondrosarcoma	0/1 ¹⁰⁷
Ewing's sarcoma	0/1 ⁸⁷
Schwannoma	0/1 ⁸⁷

*Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. No studies analyzed for c-erbB-2 mRNA or c-erbB-2 protein.

whether the protein level is increased over that of normal skin.⁵⁶ Thyroid carcinomas and adenomas can have low levels of increased c-erbB-2 mRNA. One abstract described low-level c-erbB-2 DNA amplification in one of ten salivary gland pleomorphic adenomas.⁴⁸

Correlation of c-erbB-2 Activation With Patient Outcome

Very few studies have attempted to correlate c-erbB-2 activation in non-mammary tumors with outcome. Slamon et al⁸¹ showed that c-erbB-2 amplification or overexpression in ovarian carcinomas correlates with decreased survival, especially when marked activation is present. However, they did not report the stage, histological grade, or histological subtype of these neoplasms. Another study of stages III and IV ovarian carcinomas found a correlation between decreased survival and c-erbB-2 protein overproduction, but not between survival and histological grade.¹⁸ One abstract stated that c-erbB-2 protein overproduction in 10 of 16 pulmonary adenocarcinomas correlated with decreased disease-free interval.⁷⁰ Another abstract described a tendency for immunohisto-

TABLE 13. c-erbB-2 ACTIVATION IN HUMAN TUMORS OF THE URINARY TRACT*

Tumor Type	c-erbB-2 DNA Amplification	c-erbB-2 mRNA Overproduction	c-erbB-2 Protein Overproduction
Kidney—renal cell carcinoma	1/5, ⁸⁷ 1/4, ¹⁰⁷ 0/5 ⁸⁴	0/16 ¹⁰⁴	—
Wilms' tumor	0/4 ⁸⁷	—	—
Prostate—adenocarcinoma	—	—	0/23 ⁸⁸
Urinary bladder—carcinoma	—	—	1/48 ⁸⁸

*Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. All protein studies used immunohistochemical methods.

184 T.P. SINGLETON AND J.G. STRICKLER

TABLE 14. *c-erbB-2* ACTIVATION IN MISCELLANEOUS HUMAN TUMORS*

Tumor Type	<i>c-erbB-2</i> DNA Amplification	<i>c-erbB-2</i> mRNA Overproduction	<i>c-erbB-2</i> Protein Over- production
Skin—malignant melanoma	—	—	0/10 ⁵⁸
Skin, head and neck—squamous cell carcinoma	0/7 ¹⁰⁷	—	—
Site not stated—squamous cell carcinoma	0/8, ⁵⁷ 0/2 ⁷⁵	—	—
Salivary gland—adenocarcinoma	1/1 ⁷⁵	—	—
Parotid gland—adenoid cystic carcinoma	—	—	0/1 ⁵¹
Thyroid—anaplastic carcinoma	0/1 ¹	0/1 ¹	—
Thyroid—papillary carcinoma	0/5 ¹	3(low levels)/5 ¹	—
Thyroid—adenocarcinoma	0/1 ⁵⁴	—	—
Thyroid—adenoma	0/2 ¹	1(low levels)/2 ¹	—
Neuroblastoma	0/35, ⁵¹ 0/8, ⁵⁷ 0/1 ⁷⁵	—	—
Meningioma	0/2 ⁵⁷	—	—

*Shown as number of cases with amplification (or overproduction)/total number of cases studied; reference is given as superscript. All protein studies used immunohistochemical methods.

chemical reactivity for *c-erbB-2* protein to correlate with higher grades of prostatic adenocarcinoma.⁵⁷ Additional prognostic studies of ovarian carcinomas and other neoplasms are needed.

SUMMARY

Activation of the *c-erbB-2* oncogene can occur by amplification of *c-erbB-2* DNA and by overproduction of *c-erbB-2* mRNA and *c-erbB-2* protein. Approximately 20 percent of breast carcinomas show evidence of *c-erbB-2* activation, which correlates with a poor prognosis primarily in patients with metastasis to axillary lymph nodes. Studies that have attempted to correlate *c-erbB-2* activation with other prognostic factors in breast carcinoma have reported conflicting conclusions. The pathologic and clinical significance of *c-erbB-2* activation in other neoplasms is unclear and should be assessed by additional studies.

REFERENCES

1. Aasland R, Lillehaug JR, Male R, et al. Expression of oncogenes in thyroid tumors: Coexpression of *c-erbB2/neu* and *c-erbB*. *Br J Cancer*. 57:358, 1988
2. Akiyama T, Sudo C, Ogawara H, et al. The product of the human *c-erbB-2* gene: A 185-kilodalton glycoprotein with tyrosine kinase activity. *Science*. 232:1644, 1986

3. Ali IU, Lidereau R, Theillet C, Callahan R. Reduction to homozygosity of genes on chromosome 11 in human breast neoplasia. *Science*. 238:185, 1987
4. Ali IU, Campbell G, Lidereau R, Callahan R. Amplification of c-erbB-2 and aggressive human breast tumors. *Science*. 240:1795, 1988
5. Ali IU, Campbell G, Lidereau R, Callahan R. Lack of evidence for the prognostic significance of c-erbB-2 amplification in human breast carcinoma. *Oncogene Res*. 3:139, 1988
6. Bacus SS, Bacus JW, Slamon DJ, Press MF. HER-2/neu oncogene expression and DNA ploidy analysis in breast cancer. *Arch Pathol Lab Med*. 114:164, 1990
7. Bacus SS, Ruby SG, Weinberg DS, et al. HER-2/neu oncogene expression and proliferation in breast cancers. *Am J Pathol*. 137:103, 1990
8. Bargmann CI, Hung MC, Weinberg RA. The neu oncogene encodes an epidermal growth factor receptor-related protein. *Nature*. 319:226, 1986
9. Bargmann CI, Hung MC, Weinberg RA. Multiple independent activations of the neu oncogene by a point mutation altering the transmembrane domain of p185. *Cell*. 45:649, 1986
10. Bargmann CI, Weinberg RA. Oncogenic activation of the neu-encoded receptor protein by point mutation and deletion. *EMBO J*. 7:2043, 1988
11. Barnes DM, Lammie GA, Millis RR, et al. An immunohistochemical evaluation of c-erbB-2 expression in human breast carcinoma. *Br J Cancer*. 58:448, 1988
12. Berchuck A, Kamel A, Whitaker R, et al. Overexpression of HER-2/neu is associated with poor survival in advanced epithelial ovarian cancer. *Cancer Res*. 50:4087, 1990
13. Berger MS, Locher GW, Saurer S, et al. Correlation of c-erbB-2 gene amplification and protein expression in human breast carcinoma with nodal status and nuclear grading. *Cancer Res*. 48:1238, 1988
14. Bernards R, Destree A, McKenzie S, et al. Effective tumor immunotherapy directed against an oncogene-encoded product using a vaccinia virus vector. *Proc Natl Acad Sci USA*. 84:6854, 1987
15. Bifulco I, Pozzi MR, Pierotti MA, et al. Structure and expression of oncogenes in surgical specimens of human breast carcinomas. *Br J Cancer*. 57:464, 1988
16. Borg A, Linell F, Idvall I, et al. Her2/neu amplification and comedo type breast carcinoma. *Lancet*. 1:1268, 1989
17. Borg A, Tandon AK, Sigurdsson H, et al. HER-2/neu amplification predicts poor survival in node-positive breast cancer. *Cancer Res*. 50:4332, 1990
18. Bouchard L, Lamarre L, Tremblay PJ, Jolicoeur P. Stochastic appearance of mammary tumors in transgenic mice carrying the MMTV/c-neu oncogene. *Cell*. 57:931, 1989
19. Carney WP, Retos C, Petit D, et al. Quantitation of the neu oncogene protein using a monoclonal antibody based immunoassay (abstract). *Mod Pathol*. 3:15A, 1990
20. Cline MJ, Battifora H. Abnormalities of protooncogenes in non-small cell lung cancer: Correlations with tumor type and clinical characteristics. *Cancer*. 60:2669, 1987
21. Cline MJ, Battifora H, Yokota J. Proto-oncogene abnormalities in human breast cancer: Correlations with anatomic features and clinical course of disease. *J Clin Oncol*. 5:999, 1987
22. Cohen JA, Weiner DB, More KF, et al. Expression pattern of the neu (NGL) gene-encoded growth factor receptor protein (p185^{neu}) in normal and transformed epithelial tissues of the digestive tract. *Oncogene*. 4:81, 1989
23. Colnaghi MI, Miotti S, Andreola S, et al. New prognostic factors in breast cancer (abstract). *Am Assoc Cancer Res Ann Meeting*. 30:230A, 1989

24. Coussens L, Yang-Feng TL, Liao YC, et al. Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with *neu* oncogene. *Science*. 230:1132, 1985
25. Di Fiore PP, Pierce JH, Kraus MH, et al. *erbB-2* is a potent oncogene when overexpressed in NIH/3T3 cells. *Science*. 237:178, 1987
26. Drebin JA, Link VC, Weinberg RA, Greene MI. Inhibition of tumor growth by a monoclonal antibody reactive with an oncogene-encoded tumor antigen. *Proc Natl Acad Sci USA*. 83:9129, 1986
27. Drebin JA, Link VC, Greene MI. Monoclonal antibodies reactive with distinct domains of the *neu* oncogene-encoded p185 molecule exert synergistic anti-tumor effects in vivo. *Oncogene*. 2:273, 1988
28. Drebin JA, Link VC, Greene MI. Monoclonal antibodies specific for the *neu* oncogene product directly mediate anti-tumor effects in vivo. *Oncogene*. 2:387, 1988
29. Falck VC, Gullick WJ. *c-erbB-2* oncogene product staining in gastric adenocarcinoma. An immunohistochemical study. *J Pathol*. 159:107, 1989
30. Fendly BM, Winget M, Hudziak RM, et al. Characterization of murine monoclonal antibodies reactive to either the human epidermal growth factor receptor or *HER2/neu* gene product. *Cancer Res*. 50:1550, 1990
31. Fontaine J, Tesseraux M, Klein V, et al. Gene amplification and expression of the *neu* (*c-erbB-2*) sequence in human mammary carcinoma. *Oncology*. 45:360, 1988
32. Frye RA, Benz CC, Liu E. Detection of amplified oncogenes by differential polymerase chain reaction. *Oncogene*. 4:1153, 1989
33. Fukushima SI, Matsubara KI, Yoshida M, et al. Localization of a novel *v-erbB*-related gene, *c-erbB-2*, on human chromosome 17 and its amplification in a gastric cancer cell line. *Mol Cell Biol*. 6:955, 1986
34. Guerin M, Barrois M, Terrier MJ, et al. Overexpression of either *c-myc* or *c-erbB-2/neu* proto-oncogenes in human breast carcinomas: Correlation with poor prognosis. *Oncogene Res*. 3:21, 1988
35. Guerin M, Gabillot M, Mathieu MC, et al. Structure and expression of *c-erbB-2* and EGF receptor genes in inflammatory and non-inflammatory breast cancer: Prognostic significance. *Int J Cancer*. 43:201, 1989
36. Gullick WJ, Berger MS, Bennett PLP, et al. Expression of the *c-erbB-2* protein in normal and transformed cells. *Int J Cancer*. 40:246, 1987
37. Gullick WJ, Venter DJ. The *c-erbB2* gene and its expression in human cancers. In: Waxman J, Sikora K, eds. *The Molecular Biology of Cancer*. Boston, Blackwell Sci Publ; 1989: 38-53
38. Gullick WJ. Expression of the *c-erbB-2* proto-oncogene protein in human breast cancer. *Recent Results Cancer Res*. 113:51, 1989
39. Gusterson BA, Machin LG, Gullick WJ, et al. *c-erbB-2* expression in benign and malignant breast disease. *Br J Cancer*. 58:453, 1988
40. Gusterson BA, Machin LG, Gullick WJ, et al. Immunohistochemical distribution of *c-erbB-2* in infiltrating and in situ breast cancer. *Int J Cancer*. 42:842, 1988
41. Hall PA, Hughes CM, Staddon SL, et al. The *c-erbB-2* proto-oncogene in human pancreatic cancer. *J Pathol*. 161:195, 1990
42. Hanna W, Kahn HJ, Andrulis I, Pawson T. Distribution and patterns of staining of *neu* oncogene product in benign and malignant breast diseases. *Mod Pathol*. 3:455, 1990
43. Harris AL, Nicholson S. Epidermal growth factor receptors in human breast cancer.

- In: Lippman ME, Dickson RB, eds. *Breast Cancer: Cellular and Molecular Biology*. Boston, Kluwer Academic Publ; 1988: 93-118
44. Heintz NH, Leslie KO, Rogers LA, Howard PL. Amplification of the c-erbB-2 oncogene and prognosis of breast adenocarcinoma. *Arch Pathol Lab Med.* 114:160, 1990
 45. Huettner P, Carney W, Delellis R, et al. Quantification of the neu oncogene product in ovarian neoplasms (abstract). *Mod Pathol.* 3:46A, 1990
 46. Hung MC, Yan DH, Zhao X. Amplification of the proto-neu oncogene facilitates oncogenic activation by a single point mutation. *Proc Natl Acad Sci USA.* 86:2545, 1989
 47. Hynes NE, Gerber HA, Saurer S, Groner B. Overexpression of the c-erbB-2 protein in human breast tumor cell lines. *J Cell Biochem.* 39:167, 1989
 48. Kahn HJ, Hanna W, Auger M, Andreulis I. Expression and amplification of neu oncogene in pleomorphic adenoma of salivary glands (abstract). *Mod Pathol.* 3:50A, 1990
 49. King CR, Kraus MH, Aaronson SA. Amplification of a novel v-erbB-related gene in a human mammary carcinoma. *Science.* 229:974, 1985
 50. King CR, Swain SM, Porter L, et al. Heterogeneous expression of erbB-2 messenger RNA in human breast cancer. *Cancer Res.* 49:4185, 1989
 51. Kokai Y, Dobashi K, Weiner DB, et al. Phosphorylation process induced by epidermal growth factor receptor alters the oncogenic and cellular neu (NGL) gene products. *Proc Natl Acad Sci USA.* 85:5389, 1988
 52. Lacroix H, Iglehart JD, Skinner MA, Kraus MH. Overexpression of erbB-2 or EGF receptor proteins present in early stage mammary carcinoma is detected simultaneously in matched primary tumors and regional metastases. *Oncogene.* 4:145, 1989
 53. Lemoine NR, Staddon S, Dickson C, et al. Absence of activating transmembrane mutations in the c-erbB-2 proto-oncogene in human breast cancer. *Oncogene.* 5:237, 1990
 54. Lodato RF, Maguire HC, Greene MJ, et al. Immunohistochemical evaluation of c-erbB-2 oncogene expression in ductal carcinoma in situ and atypical ductal hyperplasia of the breast. *Mod Pathol.* 3:449, 1990
 55. Maguire HC, Greene MI. The neu (c-erbB-2) oncogene. *Semin Oncol.* 16:148, 1989
 56. Maguire HC, Jaworsky C, Cohen JA, et al. Distribution of neu (c-erbB-2) protein in human skin. *J Invest Dermatol.* 89:786, 1989
 57. Masuda H, Battifora H, Yokota J, et al. Specificity of proto-oncogene amplification in human malignant diseases. *Mol Biol Med.* 4:213, 1987
 58. McCann A, Dervan PA, Johnston PA, et al. c-erbB-2 oncoprotein expression in primary human tumors. *Cancer.* 65:88, 1990
 59. McKenzie SJ, Marks PJ, Lam T, et al. Generation and characterization of monoclonal antibodies specific for the human neu oncogene product, p185. *Oncogene.* 4:543, 1989
 60. Meltzer SJ, Ahnen DJ, Battifora H, et al. Protooncogene abnormalities in colon cancers and adenomatous polyps. *Gastroenterology.* 92:1174, 1987
 61. Mori S, Akiyama T, Morishita Y, et al. Light and electron microscopical demonstration of c-erbB-2 gene product-like immunoreactivity in human malignant tumors. *Virchows Arch [B].* 54:8, 1987
 62. Mori S, Akiyama T, Yamada Y, et al. C-erbB-2 gene product, a membrane protein commonly expressed in human fetal epithelial cells. *Lab Invest.* 61:93, 1989
 63. Muller WJ, Sinn E, Pattengale PK, et al. Single-step induction of mammary

- adenocarcinoma in transgenic mice bearing the activated *c-neu* oncogene. *Cell*. 54:105, 1988
64. Ong G, Gullick W, Sikora K. Oncoprotein stability after tumor resection. *Br J Cancer*. 61:538, 1990
65. Parks HC, Lillycrop K, Howell A, Craig RK. *C-erbB2* mRNA expression in human breast tumors: Comparison with *c-erbB2* DNA amplification and correlation with prognosis. *Br J Cancer*. 61:39, 1990
66. Popescu NC, King CR, Kraus MH. Localization of the *erbB-2* gene on normal and rearranged chromosomes 17 to bands q12-21.32. *Genomics*. 4:362, 1989
67. Press MF, Pike MC, Paterson MC, et al. Overexpression of *HER-2/neu* proto-oncogene in node negative breast cancer: Correlation with increased risk of early recurrent disease (abstract). *Mod Pathol*. 3:80A, 1990
68. Ramachandra S, Machin L, Ashley S, et al. Immunohistochemical distribution of *c-erbB-2* in *in situ* breast carcinoma: A detailed morphological analysis. *J Pathol*. 161:7, 1990
69. Rio MC, Bellocq JP, Gairard B, et al. Specific expression of the *ps2* gene in subclasses of breast cancers in comparison with expression of the estrogen and progesterone receptors and the oncogene *ERBB2*. *Proc Natl Acad Sci USA*. 84:9243, 1987
70. Robinson R, Kern J, Weiner D, et al. *p185^{neu}* expression in human lung non-small cell carcinomas: An immunohistochemical study with clinicopathologic correlation (abstract). *Mod Pathol*. 3:85A, 1990
71. Rochlitz CF, Benz CC. Oncogenes in human solid tumors. In: Benz C, Liu E, eds. *Oncogenes*. Boston, Kluwer Academic Publ; 1989: 199-240
72. Sasano H, Garret CT, Wilkinson DS, et al. Protooncogene amplification and tumor ploidy in human ovarian neoplasms. *Hum Pathol*. 21:382, 1990
73. Schechter AL, Stern DF, Vaidyanathan L, et al. The *neu* oncogene: An *erbB*-related gene encoding a 185,000-M_r tumor antigen. *Nature*. 312:513, 1984
74. Schechter AL, Hung MC, Vaidyanathan L, et al. The *neu* gene: An *erbB*-homologous gene distinct from and unlinked to the gene encoding the EGF receptor. *Science*. 229:976, 1985
75. Schneider PM, Hung MC, Chiocca SM, et al. Differential expression of the *c-erbB-2* gene in human small cell and non-small cell lung cancer. *Cancer Res*. 49:4968, 1989
76. Semba-K, Kamata N, Toyoshima K, Yamamoto T. A *v-erbB*-related protooncogene, *c-erbB-2*, is distinct from the *c-erbB-1*/epidermal growth factor-receptor gene and is amplified in a human salivary gland adenocarcinoma. *Proc Natl Acad Sci USA*. 82:6497, 1985
77. Seshadri R, Matthews C, Dobrovic A, Horsfall DJ. The significance of oncogene amplification in primary breast cancer. *Int J Cancer*. 43:270, 1989
78. Shih C, Padhy LC, Murray M, Weinberg RA. Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts. *Nature*. 290:261, 1981
79. Slamon DJ, Clark GM, Wong SG, et al. Human breast cancer: Correlation of relapse and survival with amplification of the *HER-2/neu* oncogene. *Science*. 235:177, 1987
80. Slamon DJ, Clark GM. Amplification of *c-erbB-2* and aggressive human breast tumors. *Science*. 240:1795, 1988
81. Slamon DJ, Godolphin W, Jones LA, et al. Studies of the *HER-2/neu* proto-oncogene in human breast and ovarian cancer. *Science*. 244:707, 1989

82. Slebos RJC, Evers SG, Wagenaar SS, Rodenhuis S. Cellular protooncogenes are infrequently amplified in untreated non-small cell lung cancer. *Br J Cancer*. 59:76, 1989
83. Stern DF, Kamps MP, Cao H. Oncogenic activation of p185^{neu} stimulates tyrosine phosphorylation in vivo. *Mol Cell Biol*. 8:3969, 1988
84. Tal M, Wetzler M, Josephberg Z, et al. Sporadic amplification of the *HER2/neu* protooncogene in adenocarcinomas of various tissues. *Cancer Res*. 48:1517, 1988
85. Tandon AK, Clark GM, Chamness CC, et al. *HER-2/neu* oncogene protein and prognosis in breast cancer. *J Clin Oncol*. 7:1120, 1989
86. Thor AD, Schwartz LH, Koerner FC, et al. Analysis of c-erbB-2 expression in breast carcinomas with clinical follow-up. *Cancer Res*. 49:7147, 1989
87. Tsuda H, Hirohashi S, Shimamoto Y, et al. Correlation between long-term survival in breast cancer patients and amplification of two putative oncogene-coamplification units: *hst-1/int-2* and c-erbB-2/*lar-1*. *Cancer Res*. 49:3104, 1989
88. Tsutsumi Y, Naber SP, DeLellis RA, et al. *Neu* oncogene protein and epidermal growth factor receptor are independently expressed in benign and malignant breast tissues. *Hum Pathol*. 21:750, 1990
89. Tsutsumi Y, Stork PJ, Wolfe HJ. Detection of DNA amplification and mRNA overexpression of the *neu* oncogene in breast carcinomas by polymerase chain reaction (abstract). *Mod Pathol*. 3:101A, 1990
90. Van de Vijver M, van de Berselaar R, Devilee P, et al. Amplification of the *neu* (c-erbB-2) oncogene in human mammary tumors is relatively frequent and is often accompanied by amplification of the linked c-erbA oncogene. *Mol Cell Biol*. 7:2019, 1987
91. Van de Vijver MJ, Mooi WJ, Wisman P, et al. Immunohistochemical detection of the *neu* protein in tissue sections of human breast tumors with amplified *neu* DNA. *Oncogene*. 2:175, 1988
92. Van de Vijver MJ, Peterse JL, Mooi WJ, et al. *Neu*-protein overexpression in breast cancer: Association with comedo-type ductal carcinoma in situ and limited prognostic value in stage II breast cancer. *N Engl J Med*. 319:1239, 1988
93. Varley JM, Swallow JE, Brammar WJ, et al. Alterations to either c-erbB-2 (*neu*) or c-*myc* proto-oncogenes in breast carcinomas correlate with poor short-term prognosis. *Oncogene*. 1:423, 1987
94. Venter DJ, Tuzi NL, Kumar S, Gullick WJ. Overexpression of the c-erbB-2 oncoprotein in human breast carcinomas: Immunohistological assessment correlates with gene amplification. *Lancet*. 2:69, 1987
95. Voravud N, Foster CS, Gilbertson JA, et al. Oncogene expression in cholangiocarcinoma and in normal hepatic development. *Hum Pathol*. 20:1163, 1989
96. Walker RA, Senior PV, Jones JL, et al. An immunohistochemical and in situ hybridization study of c-*myc* and c-erbB-2 expression in primary human breast carcinomas. *J Pathol*. 158:97, 1989
97. Ware JL, Maygarden SJ, Koontz WW, Strom SC. Differential reactivity with anti-c-erbB-2 antiserum among human malignant and benign prostatic tissue (abstract). *Am Assoc Cancer Res Ann Meeting*. 30:437A, 1989
98. Weiner DB, Liu J, Cohen JA, et al. A point mutation in the *neu* oncogene mimics ligand induction of receptor aggregation. *Nature*. 339:230, 1989
99. Weiner DB, Nordberg J, Robinson R, et al. Expression of the *neu* gene-encoded protein (p185^{neu}) in human non-small cell carcinomas of the lung. *Cancer Res*. 50:421, 1990

190 T.P. SINGLETON AND J.G. STRICKLER

100. Wells A. The epidermal growth factor receptor and its ligand. In: Benz C, Liu E, eds. *Oncogenes*. Boston, Kluwer Academic Pub; 1989: 143-168
101. Wright C, Angus B, Nicholson S, et al. Expression of c-erbB-2 oncoprotein: A prognostic indicator in human breast cancer. *Cancer Res.* 49:2087, 1989
102. Wu A, Colombero A, Low J, et al. Analysis of expression and mutation of the erbB-2 gene in breast carcinoma by the polymerase chain reaction (abstract). *Mod Pathol.* 3:108A, 1990
103. Yamamoto T, Ikawa S, Akiyama T, et al. Similarity of protein encoded by the human c-erbB-2 gene to epidermal growth factor receptor. *Nature.* 319:230, 1986
104. Yao M, Shuin T, Miski H, Kubota Y. Enhanced expression of c-myc and epidermal growth factor receptor (C-erbB-1) genes in primary human renal cancer. *Cancer Res.* 48:6753, 1988
105. Yarden Y, Weinberg RA. Experimental approaches to hypothetical hormones: Detection of a candidate ligand of the neu protooncogene. *Proc Natl Acad Sci USA.* 86:3179, 1989
106. Yee LD, Kacinski BM, Carter D. Oncogene structure, function and expression in breast cancer. *Semin Diagn Pathol.* 6:110, 1989
107. Yokota J, Yamamoto T, Toyoshima K, et al. Amplification of c-erbB-2 oncogene in human adenocarcinomas in vivo. *Lancet.* 1:765, 1986
108. Yokota J, Yamamoto T, Miyajima N, et al. Genetic alterations of the c-erbB-2 oncogene occur frequently in tubular adenocarcinoma of the stomach and are often accompanied by amplification of the v-erbA homologue. *Oncogene.* 2:283, 1988
109. Zeillinger R, Kury F, Czerwenka K, et al. HER-2 amplification, steroid receptors and epidermal growth factor receptor in primary breast cancer. *Oncogene.* 4:109, 1989
110. Zhang X, Silva E, Gershenson D, Hung MC. Amplification and rearrangement of c-erbB proto-oncogenes in cancer of human female genital tract. *Oncogene.* 4:985, 1989
111. Zhou D, Battifora H, Yokota J, et al. Association of multiple copies of the c-erbB-2 oncogene with spread of breast cancer. *Cancer Res.* 47:6123, 1987
112. Zhou D, Gonzalez-Cadavid N, Ahuja H, et al. A unique pattern of proto-oncogene abnormalities in ovarian adenocarcinomas. *Cancer.* 62:1573, 1988
113. Zhou D, Ahuja H, Cline MJ. Proto-oncogene abnormalities in human breast cancer: c-erbB-2 amplification does not correlate with recurrence of disease. *Oncogene.* 4:105, 1989

DECLARATION OF PAUL POLAKIS, Ph.D.

I, Paul Polakis, Ph.D., declare and say as follows:

1. I was awarded a Ph.D. by the Department of Biochemistry of the Michigan State University in 1984. My scientific Curriculum Vitae is attached to and forms part of this Declaration (Exhibit A).
2. I am currently employed by Genentech, Inc. where my job title is Staff Scientist. Since joining Genentech in 1999, one of my primary responsibilities has been leading Genentech's Tumor Antigen Project, which is a large research project with a primary focus on identifying tumor cell markers that find use as targets for both the diagnosis and treatment of cancer in humans.
3. As part of the Tumor Antigen Project, my laboratory has been analyzing differential expression of various genes in tumor cells relative to normal cells. The purpose of this research is to identify proteins that are abundantly expressed on certain tumor cells and that are either (i) not expressed, or (ii) expressed at lower levels, on corresponding normal cells. We call such differentially expressed proteins "tumor antigen proteins". When such a tumor antigen protein is identified, one can produce an antibody that recognizes and binds to that protein. Such an antibody finds use in the diagnosis of human cancer and may ultimately serve as an effective therapeutic in the treatment of human cancer.
4. In the course of the research conducted by Genentech's Tumor Antigen Project, we have employed a variety of scientific techniques for detecting and studying differential gene expression in human tumor cells relative to normal cells, at genomic DNA, mRNA and protein levels. An important example of one such technique is the well known and widely used technique of microarray analysis which has proven to be extremely useful for the identification of mRNA molecules that are differentially expressed in one tissue or cell type relative to another. In the course of our research using microarray analysis, we have identified approximately 200 gene transcripts that are present in human tumor cells at significantly higher levels than in corresponding normal human cells. To date, we have generated antibodies that bind to about 30 of the tumor antigen proteins expressed from these differentially expressed gene transcripts and have used these antibodies to quantitatively determine the level of production of these tumor antigen proteins in both human cancer cells and corresponding normal cells. We have then compared the levels of mRNA and protein in both the tumor and normal cells analyzed.
5. From the mRNA and protein expression analyses described in paragraph 4 above, we have observed that there is a strong correlation between changes in the level of mRNA present in any particular cell type and the level of protein

expressed from that mRNA in that cell type. In approximately 80% of our observations we have found that increases in the level of a particular mRNA correlates with changes in the level of protein expressed from that mRNA when human tumor cells are compared with their corresponding normal cells.

6. Based upon my own experience accumulated in more than 20 years of research, including the data discussed in paragraphs 4 and 5 above and my knowledge of the relevant scientific literature, it is my considered scientific opinion that for human genes, an increased level of mRNA in a tumor cell relative to a normal cell typically correlates to a similar increase in abundance of the encoded protein in the tumor cell relative to the normal cell. In fact, it remains a central dogma in molecular biology that increased mRNA levels are predictive of corresponding increased levels of the encoded protein. While there have been published reports of genes for which such a correlation does not exist, it is my opinion that such reports are exceptions to the commonly understood general rule that increased mRNA levels are predictive of corresponding increased levels of the encoded protein.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

Dated: 5/07/04

By: Paul Polakis
Paul Polakis, Ph.D.

CURRICULUM VITAE

PAUL G. POLAKIS
Staff Scientist
Genentech, Inc
1 DNA Way, MS#40
S. San Francisco, CA 94080

EDUCATION:

Ph.D., Biochemistry, Department of Biochemistry,
Michigan State University (1984)

B.S., Biology. College of Natural Science, Michigan State University (1977)

PROFESSIONAL EXPERIENCE:

2002-present

Staff Scientist, Genentech, Inc
S. San Francisco, CA

1999- 2002

Senior Scientist, Genentech, Inc.,
S. San Francisco, CA

1997 -1999

Research Director
Onyx Pharmaceuticals, Richmond, CA

1992- 1996

Senior Scientist, Project Leader, Onyx
Pharmaceuticals, Richmond, CA

1991-1992

Senior Scientist, Chiron Corporation,
Emeryville, CA.

1989-1991

Scientist, Cetus Corporation, Emeryville CA.

1987-1989

Postdoctoral Research Associate, Genentech,
Inc., South San Francisco, CA.

1985-1987

Postdoctoral Research Associate, Department
of Medicine, Duke University Medical Center,
Durham, NC

1984-1985

Assistant Professor, Department of Chemistry,
Oberlin College, Oberlin, Ohio

1980-1984

Graduate Research Assistant, Department of
Biochemistry, Michigan State University
East Lansing, Michigan

PUBLICATIONS:

1. Polakis, P. G. and Wilson, J. E. 1982 Purification of a Highly Bindable Rat Brain Hexokinase by High Performance Liquid Chromatography. **Biochem. Biophys. Res. Commun.** 107, 937-943.
2. Polakis, P. G. and Wilson, J. E. 1984 Proteolytic Dissection of Rat Brain Hexokinase: Determination of the Cleavage Pattern during Limited Digestion with Trypsin. **Arch. Biochem. Biophys.** 234, 341-352.
3. Polakis, P. G. and Wilson, J. E. 1985 An Intact Hydrophobic N-Terminal Sequence is Required for the Binding of Rat Brain Hexokinase to Mitochondria. **Arch. Biochem. Biophys.** 236, 328-337.
4. Uhing, R. J., Polakis, P. G. and Snyderman, R. 1987 Isolation of GTP-binding Proteins from Myeloid HL60 Cells. **J. Biol. Chem.** 262, 15575-15579.
5. Polakis, P. G., Uhing, R. J. and Snyderman, R. 1988 The Formylpeptide Chemoattractant Receptor Copurifies with a GTP-binding Protein Containing a Distinct 40 kDa Pertussis Toxin Substrate. **J. Biol. Chem.** 263, 4969-4979.
6. Uhing, R. J., Dillon, S., Polakis, P. G., Truett, A. P. and Snyderman, R. 1988 Chemoattractant Receptors and Signal Transduction Processes in Cellular and Molecular Aspects of Inflammation (Poste, G. and Crooke, S. T. eds.) pp 335-379.
7. Polakis, P. G., Evans, T. and Snyderman 1989 Multiple Chromatographic Forms of the Formylpeptide Chemoattractant Receptor and their Relationship to GTP-binding Proteins. **Biochem. Biophys. Res. Commun.** 161, 276-283.
8. Polakis, P. G., Snyderman, R. and Evans, T. 1989 Characterization of G25K, a GTP-binding Protein Containing a Novel Putative Nucleotide Binding Domain. **Biochem. Biophys. Res. Commun.** 160, 25-32.
9. Polakis, P., Weber, R. F., Nevins, B., Didsbury, J., Evans, T. and Snyderman, R. 1989 Identification of the *ral* and *rac1* Gene Products, Low Molecular Mass GTP-binding Proteins from Human Platelets. **J. Biol. Chem.** 264, 16383-16389.
10. Snyderman, R., Perianin, A., Evans, T., Polakis, P. and Didsbury, J. 1989 G Proteins and Neutrophil Function. In ADP-Ribosylating Toxins and G Proteins: Insights into Signal Transduction. (J. Moss and M. Vaughn, eds.) Amer. Soc. Microbiol. pp. 295-323.

11. Hart, M.J., Polakis, P.G., Evans, T. and Cerrione, R.A. 1990 The Identification and Characterization of an Epidermal Growth Factor-Stimulated Phosphorylation of a Specific Low Molecular Mass GTP-binding Protein in a Reconstituted Phospholipid Vesicle System. *J. Biol. Chem.* 265, 5990-6001.
12. Yatani, A., Okabe, K., Polakis, P., Halenbeck, R., McCormick, F. and Brown, A. M. 1990 ras p21 and GAP Inhibit Coupling of Muscarinic Receptors to Atrial K⁺ Channels. *Cell*. 61, 769-776.
13. Munemitsu, S., Innis, M.A., Clark, R., McCormick, F., Ullrich, A. and Polakis, P.G. 1990 Molecular Cloning and Expression of a G25K cDNA, the Human Homolog of the Yeast Cell Cycle Gene CDC42. *Mol. Cell. Biol.* 10, 5977-5982.
14. Polakis, P.G., Rubinfeld, B., Evans, T. and McCormick, F. 1991 Purification of Plasma Membrane-Associated GTPase Activating Protein (GAP) Specific for rap-1/krev-1 from HL60 Cells. *Proc. Natl. Acad. Sci. USA* 88, 239-243.
15. Moran, M. F., Polakis, P., McCormick, F., Pawson, T. and Ellis, C. 1991 Protein Tyrosine Kinases Regulate the Phosphorylation, Protein Interactions, Subcellular Distribution, and Activity of p21ras GTPase Activating Protein. *Mol. Cell. Biol.* 11, 1804-1812.
16. Rubinfeld, B., Wong, G., Bekesi, E., Wood, A., McCormick, F. and Polakis, P. G. 1991 A Synthetic Peptide Corresponding to a Sequence in the GTPase Activating Protein Inhibits p21^{ras} Stimulation and Promotes Guanine Nucleotide Exchange. *Internatl. J. Peptide and Prot. Res.* 38, 47-53.
17. Rubinfeld, B., Munemitsu, S., Clark, R., Conroy, L., Watt, K., Crosier, W., McCormick, F., and Polakis, P. 1991 Molecular Cloning of a GTPase Activating Protein Specific for the Krev-1 Protein p21^{rap1}. *Cell* 65, 1033-1042.
18. Zhang, K., Papageorge, A., G., Martin, P., Vass, W. C., Olah, Z., Polakis, P., McCormick, F. and Lowy, D. R. 1991 Heterogenous Amino Acids in RAS and Rap1A Specifying Sensitivity to GAP Proteins. *Science* 254, 1630-1634.
19. Martin, G., Yatani, A., Clark, R., Polakis, P., Brown, A. M. and McCormick, F. 1992 GAP Domains Responsible for p21^{ras}-dependent Inhibition of Muscarinic Atrial K⁺ Channel Currents. *Science* 255, 192-194.
20. McCormick, F., Martin, G. A., Clark, R., Bollag, G. and Polakis, P. 1992 Regulation of p21^{ras} by GTPase Activating Proteins. *Cold Spring Harbor Symposia on Quantitative Biology*. Vol. 56, 237-241.
21. Pronk, G. B., Polakis, P., Wong, G., deVries-Smits, A. M., Bos J. L. and McCormick, F. 1992 p60^{v-src} Can Associate with and Phosphorylate the p21^{ras} GTPase Activating Protein. *Oncogene* 7,389-394.
22. Polakis P. and McCormick, F. 1992 Interactions Between p21^{ras} Proteins and Their GTPase Activating Proteins. In Cancer Surveys (Franks, L. M., ed.) 12, 25-42.

23. Wong, G., Muller, O., Clark, R., Conroy, L., Moran, M., Polakis, P. and McCormick, F. 1992 Molecular cloning and nucleic acid binding properties of the GAP-associated tyrosine phosphoprotein p62. *Cell* 69, 551-558.

24. Polakis, P., Rubinfeld, B. and McCormick, F. 1992 Phosphorylation of rap1GAP in vivo and by cAMP-dependent Kinase and the Cell Cycle p34^{cdc2} Kinase in vitro. *J. Biol. Chem.* 267, 10780-10785.

25. McCabe, P.C., Haubrauck, H., Polakis, P., McCormick, F., and Innis, M. A. 1992 Functional Interactions Between p21^{rap1A} and Components of the Budding pathway of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 12, 4084-4092.

26. Rubinfeld, B., Crosier, W.J., Albert, I., Conroy, L., Clark, R., McCormick, F. and Polakis, P. 1992 Localization of the rap1GAP Catalytic Domain and Sites of Phosphorylation by Mutational Analysis. *Mol. Cell. Biol.* 12, 4634-4642.

27. Ando, S., Kaibuchi, K., Sasaki, K., Hiraoka, T., Nishiyama, T., Mizuno, T., Asada, M., Nunoi, H., Matsuda, I., Matsuura, Y., Polakis, P., McCormick, F. and Takai, Y. 1992 Post-translational processing of rac p21s is important both for their interaction with the GDP/GTP exchange proteins and for their activation of NADPH oxidase. *J. Biol. Chem.* 267, 25709-25713.

28. Janoueix-Lerosey, I., Polakis, P., Tavittian, A. and deGunzberg, J. 1992 Regulation of the GTPase activity of the ras-related rap2 protein. *Biochem. Biophys. Res. Commun.* 189, 455-464.

29. Polakis, P. 1993 GAPs Specific for the rap1/Krev-1 Protein. in GTP-binding Proteins: the ras-superfamily. (J.C. LaCale and F. McCormick, eds.) 445-452.

30. Polakis, P. and McCormick, F. 1993 Structural requirements for the interaction of p21^{ras} with GAP, exchange factors, and its biological effector target. *J. Biol. Chem.* 268, 9157-9160.

31. Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S., Masiarz, F., Munemitsu, S. and Polakis, P. 1993 Association of the APC gene product with beta- catenin. *Science* 262, 1731-1734.

32. Weiss, J., Rubinfeld, B., Polakis, P., McCormick, F. Cavenee, W. A. and Arden, K. 1993 The gene for human rap1-GTPase activating protein (rap1GAP) maps to chromosome 1p35-1p36.1. *Cytogenet. Cell Genet.* 66, 18-21.

33. Sato, K. Y., Polakis, P., Haubruck, H., Fasching, C. L., McCormick, F. and Stanbridge, E. J. 1994 Analysis of the tumor suppressor activity of the K-rev gene in human tumor cell lines. *Cancer Res.* 54, 552-559.

34. Janoueix-Lerosey, I., Fontenay, M., Tobelem, G., Tavittian, A., Polakis, P. and DeGunzburg, J. 1994 Phosphorylation of rap1GAP during the cell cycle. *Biochem. Biophys. Res. Commun.* 202, 967-975

35. Munemitsu, S., Souza, B., Mueller, O., Albert, I., Rubinfeld, B., and Polakis, P. 1994 The APC gene product associates with microtubules in vivo and affects their assembly in vitro. *Cancer Res.* 54, 3676-3681.

36. Rubinfeld, B. and Polakis, P. 1995 Purification of baculovirus produced rap1GAP. **Methods Enz.** 255,31
37. Polakis, P. 1995 Mutations in the APC gene and their implications for protein structure and function. **Current Opinions In Genetics and Development** 5, 66-71
38. Rubinfeld, B., Souza, B., Albert, I., Munemitsu, S. and Polakis P. 1995 The APC protein and E-cadherin form similar but independent complexes with α -catenin, β -catenin and Plakoglobin. **J. Biol. Chem.** 270, 5549-5555
39. Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and Polakis, P. 1995 Regulation of intracellular β -catenin levels by the APC tumor suppressor gene. **Proc. Natl. Acad. Sci.** 92, 3046-3050.
40. Lock, P., Fumagalli, S., Polakis, P. McCormick, F. and Courtneidge, S. A. 1996 The human p62 cDNA encodes Sam68 and not the rasGAP-associated p62 protein. **Cell** 84, 23-24.
41. Papkoff, J., Rubinfeld, B., Schryver, B. and Polakis, P. 1996 Wnt-1 regulates free pools of catenins and stabilizes APC-catenin complexes. **Mol. Cell. Biol.** 16, 2128-2134.
42. Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S. and Polakis, P. 1996 Binding of GSK3 β to the APC- β -catenin complex and regulation of complex assembly. **Science** 272, 1023-1026.
43. Munemitsu, S., Albert, I., Rubinfeld, B. and Polakis, P. 1996 Deletion of amino-terminal structure stabilizes β -catenin in vivo and promotes the hyperphosphorylation of the APC tumor suppressor protein. **Mol. Cell. Biol.** 16, 4088-4094.
44. Hart, M. J., Callow, M. G., Sousa, B. and Polakis P. 1996 IQGAP1, a calmodulin binding protein with a rasGAP related domain, is a potential effector for cdc42Hs. **EMBO J.** 15, 2997-3005.
45. Nathke, I. S., Adams, C. L., Polakis, P., Sellin, J. and Nelson, W. J. 1996 The adenomatous polyposis coli (APC) tumor suppressor protein is localized to plasma membrane sites involved in active epithelial cell migration. **J. Cell. Biol.** 134, 165-180.
46. Hart, M. J., Sharma, S., elMasry, N., Qui, R-G., McCabe, P., Polakis, P. and Bollag, G. 1996 Identification of a novel guanine nucleotide exchange factor for the rho GTPase. **J. Biol. Chem.** 271, 25452.
47. Thomas JE, Smith M, Rubinfeld B, Gutowski M, Beckmann RP, and Polakis P. 1996 Subcellular localization and analysis of apparent 180-kDa and 220-kDa proteins of the breast cancer susceptibility gene, BRCA1. **J. Biol. Chem.** 1996 271, 28630-28635
48. Hayashi, S., Rubinfeld, B., Souza, B., Polakis, P., Wieschaus, E., and Levine, A. 1997 A Drosophila homolog of the tumor suppressor adenomatous polyposis coli

down-regulates β -catenin but its zygotic expression is not essential for the regulation of armadillo. *Proc. Natl. Acad. Sci.* 94, 242-247.

49. Vleminckx, K., Rubinfeld, B., Polakis, P. and Gumbiner, B. 1997 The APC tumor suppressor protein induces a new axis in *Xenopus* embryos. *J. Cell. Biol.* 136, 411-420.

50. Rubinfeld, B., Robbins, P., El-Gamil, M., Albert, I., Porfiri, P. and Polakis, P. 1997 Stabilization of β -catenin by genetic defects in melanoma cell lines. *Science* 275, 1790-1792.

51. Polakis, P. The adenomatous polyposis coli (APC) tumor suppressor. 1997 *Biochem. Biophys. Acta*, 1332, F127-F147.

52. Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and Polakis, P. 1997 Loss of β -catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. *Cancer Res.* 57, 4624-4630.

53. Porfiri, E., Rubinfeld, B., Albert, I., Hovanes, K., Waterman, M., and Polakis, P. 1997 Induction of a β -catenin-LEF-1 complex by wnt-1 and transforming mutants of β -catenin. *Oncogene* 15, 2833-2839.

54. Thomas JE, Smith M, Tonkinson JL, Rubinfeld B, and Polakis P., 1997 Induction of phosphorylation on BRCA1 during the cell cycle and after DNA damage. *Cell Growth Differ.* 8, 801-809.

55. Hart, M., de los Santos, R., Albert, I., Rubinfeld, B., and Polakis P., 1998 Down regulation of β -catenin by human Axin and its association with the adenomatous polyposis coli (APC) tumor suppressor, β -catenin and glycogen synthase kinase 3 β . *Current Biology* 8, 573-581.

56. Polakis, P. 1998 The oncogenic activation of β -catenin. *Current Opinions in Genetics and Development* 9, 15-21

57. Matt Hart, Jean-Paul Concordet, Irina Lassot, Iris Albert, Rico del los Santos, Hervé Durand, Christine Perret, Bonnee Rubinfeld, Florence Margottin, Richard Benarous and Paul Polakis. 1999 The F-box protein β -TrCP associates with phosphorylated β -catenin and regulates its activity in the cell. *Current Biology* 9, 207-10.

58. Howard C. Crawford, Barbara M. Fingleton, Bonnee Rubinfeld, Paul Polakis and Lynn M. Matrisian. 1999 The metalloproteinase matrilysin is a target of β -catenin transactivation in intestinal tumours. *Oncogene* 18, 2883-91.

59. Meng J, Glick JL, Polakis P, Casey PJ. 1999 Functional interaction between Galpha(z) and Rap1GAP suggests a novel form of cellular cross-talk. *J Biol Chem.* 17, 36663-9

60. Vijayasurian Easwaran, Virginia Song, Paul Polakis and Steve Byers 1999 The ubiquitin-proteasome pathway and serine kinase activity modulate APC mediated regulation of β -catenin-LEF signaling. **J. Biol. Chem.** 274(23):16641-5.
61. Polakis P, Hart M and Rubinfeld B. 1999 Defects in the regulation of beta-catenin in colorectal cancer. **Adv Exp Med Biol.** 470, 23-32
62. Shen Z, Batzer A, Koehler JA, Polakis P, Schlessinger J, Lydon NB, Moran MF. 1999 Evidence for SH3 domain directed binding and phosphorylation of Sam68 by Src. **Oncogene.** 18, 4647-53
64. Thomas GM, Frame S, Goedert M, Nathke I, Polakis P, Cohen P. 1999 A GSK3- binding peptide from FRAT1 selectively inhibits the GSK3-catalysed phosphorylation of axin and beta-catenin. **FEBS Lett.** 458, 247-51.
65. Peifer M, Polakis P. 2000 Wnt signaling in oncogenesis and embryogenesis—a look outside the nucleus. **Science** 287,1606-9.
66. Polakis P. 2000 Wnt signaling and cancer. **Genes Dev**;14, 1837-1851.
67. Spink KE, Polakis P, Weis WI 2000 Structural basis of the Axin-adenomatous polyposis coli interaction. **EMBO J** 19, 2270-2279.
68. Szeto, W., Jiang, W., Tice, D.A., Rubinfeld, B., Hollingshead, P.G., Fong, S.E., Dugger, D.L., Pham, T., Yansura, D.E., Wong, T.A., Grimaldi, J.C., Corpuz, R.T., Singh J.S., Frantz, G.D., Devaux, B., Crowley, C.W., Schwall, R.H., Eberhard, D.A., Rastelli, L., Polakis, P. and Pennica, D. 2001 Overexpression of the Retinoic Acid-Responsive Gene Stra6 in Human Cancers and its Synergistic Induction by Wnt-1 and Retinoic Acid. **Cancer Res** 61, 4197-4204.
69. Rubinfeld B, Tice DA, Polakis P. 2001 Axin dependent phosphorylation of the adenomatous polyposis coli protein mediated by casein kinase 1 epsilon. **J Biol Chem** 276, 39037-39045.
70. Polakis P. 2001 More than one way to skin a catenin. **Cell** 2001 105, 563-566.
71. Tice DA, Soloviev I, Polakis P. 2002 Activation of the Wnt Pathway Interferes with Serum Response Element-driven Transcription of Immediate Early Genes. **J Biol Chem.** 277, 6118-6123.
72. Tice DA, Szeto W, Soloviev I, Rubinfeld B, Fong SE, Dugger DL, Winer J,

Williams PM, Wieand D, Smith V, Schwall RH, Pennica D, Polakis P. 2002 Synergistic activation of tumor antigens by wnt-1 signaling and retinoic acid revealed by gene expression profiling. *J Biol Chem.* 277,14329-14335.

73. Polakis, P. 2002 Casein kinase I: A wnt'er of disconnect. *Curr. Biol.* 12, R499.

74. Mao, W., Luis, E., Ross, S., Silva, J., Tan, C., Crowley, C., Chui, C., Franz, G., Senter, P., Koeppen, H., Polakis, P. 2004 EphB2 as a therapeutic antibody drug target for the treatment of colorectal cancer. *Cancer Res.* 64, 781-788.

75. Shibamoto, S., Winer, J., Williams, M., Polakis, P. 2003 A Blockade in Wnt signaling is activated following the differentiation of F9 teratocarcinoma cells. *Exp. Cell Res.* 29211-20.

76. Zhang Y, Eberhard DA, Frantz GD, Dowd P, Wu TD, Zhou Y, Watanabe C, Luoh SM, Polakis P, Hillan KJ, Wood WI, Zhang Z. 2004 GEPIS--quantitative gene expression profiling in normal and cancer tissues. *Bioinformatics*, April 8

MOLECULAR BIOLOGY OF
THE CELL
THIRD EDITION

Text Editor: Miranda Robertson
Managing Editor: Ruth Adams
Illustrator: Nigel Orme
Molecular Model Drawings: Kate Hesketh-Moore
Director of Electronic Publishing: John M-Roblin
Computer Specialist: Chuck Bartelt
Disk Preparation: Carol Winter
Copy Editor: Shirley M. Cobert
Production Editor: Douglas Goertzen
Production Coordinator: Perry Bessas
Indexer: Maija Hinkle

Bruce Alberts received his Ph.D. from Harvard University and is currently President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. *Dennis Bray* received his Ph.D. from the Massachusetts Institute of Technology and is currently a Medical Research Council Fellow in the Department of Zoology, University of Cambridge. *Julian Lewis* received his D.Phil. from the University of Oxford and is currently a Senior Scientist in the Imperial Cancer Research Fund Developmental Biology Unit, University of Oxford. *Martin Raff* received his M.D. from McGill University and is currently a Professor in the MRC Laboratory for Molecular Cell Biology and the Biology Department, University College London. *Keith Roberts* received his Ph.D. from the University of Cambridge and is currently Head of the Department of Cell Biology, the John Innes Institute, Norwich. *James D. Watson* received his Ph.D. from Indiana University and is currently Director of the Cold Spring Harbor Laboratory. He is the author of *Molecular Biology of the Gene* and, with Francis Crick and Maurice Wilkins, won the Nobel Prize in Medicine and Physiology in 1962.

© 1983, 1989, 1994 by Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson.

All rights reserved. No part of this book covered by the copyright hereon may be reproduced or used in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the publisher.

Library of Congress Cataloging-in-Publication Data

Molecular biology of the cell / Bruce Alberts . . . [et al.].—3rd ed.

p. cm.

Includes bibliographical references and index.

ISBN 0-8153-1619-4 (hard cover).—ISBN 0-8153-1620-8 (pbk.)

1. Cytology. 2. Molecular biology. I. Alberts, Bruce.

[DNLM: 1. Cells. 2. Molecular Biology. QH 581.2 M718 1994]

QH581.2.M64 1994

574.87—dc20

DNLM/DLC

for Library of Congress

93-45907

CIP

Published by Garland Publishing, Inc.
717 Fifth Avenue, New York, NY 10022

Printed in the United States of America

15 14 13 12 10 9 8 7

Front cover: The photograph shows a rat nerve cell in culture. It is labeled (*yellow*) with a fluorescent antibody that stains its cell body and dendritic processes. Nerve terminals (*green*) from other neurons (not visible), which have made synapses on the cell, are labeled with a different antibody. (Courtesy of Olaf Mundigl and Pietro de Camilli.)

Dedication page: Gavin Borden, late president of Garland Publishing, weathered in during his mid-1980s climb near Mount McKinley with MBoC author Bruce Alberts and famous mountaineer guide Mugs Stump (1940–1992).

Back cover: The authors, in alphabetical order, crossing Abbey Road in London on their way to lunch. Much of this third edition was written in a house just around the corner. (Photograph by Richard Olivier.)

extracts. If these minor cell proteins differ among cells to the same extent as the more abundant proteins, as is commonly assumed, only a small number of protein differences (perhaps several hundred) suffice to create very large differences in cell morphology and behavior.

A Cell Can Change the Expression of Its Genes in Response to External Signals³

Most of the specialized cells in a multicellular organism are capable of altering their patterns of gene expression in response to extracellular cues. If a liver cell is exposed to a glucocorticoid hormone, for example, the production of several specific proteins is dramatically increased. Glucocorticoids are released during periods of starvation or intense exercise and signal the liver to increase the production of glucose from amino acids and other small molecules; the set of proteins whose production is induced includes enzymes such as tyrosine aminotransferase, which helps to convert tyrosine to glucose. When the hormone is no longer present, the production of these proteins drops to its normal level.

Other cell types respond to glucocorticoids in different ways. In fat cells, for example, the production of tyrosine aminotransferase is reduced, while some other cell types do not respond to glucocorticoids at all. These examples illustrate a general feature of cell specialization—different cell types often respond in different ways to the same extracellular signal. Underlying this specialization are features that do not change, which give each cell type its permanently distinctive character. These features reflect the persistent expression of different sets of genes.

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein⁴

If differences between the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? There are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the primary RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytoplasm (**RNA transport control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 9-2).

For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 9-2, only transcriptional control ensures that no superfluous intermediates are synthesized. In the

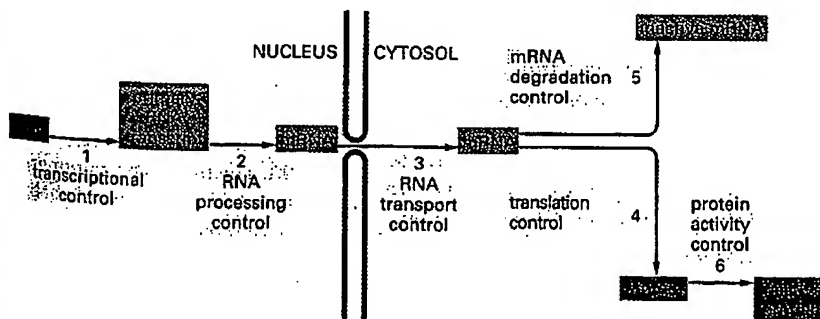


Figure 9-2 Six steps at which eucaryote gene expression can be controlled. Only controls that operate at steps 1 through 5 are discussed in this chapter. The regulation of protein activity (step 6) is discussed in Chapter 5; this includes reversible activation or inactivation by protein phosphorylation as well as irreversible inactivation by proteolytic degradation.

following sections we discuss the DNA and protein components that regulate the initiation of gene transcription. We return at the end of the chapter to the other ways of regulating gene expression.

Summary

The genome of a cell contains in its DNA sequence the information to make many thousands of different protein and RNA molecules. A cell typically expresses only a fraction of its genes, and the different types of cells in multicellular organisms arise because different sets of genes are expressed. Moreover, cells can change the pattern of genes they express in response to changes in their environment, such as signals from other cells. Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of RNA transcription is the most important point of control.

DNA-binding Motifs in Gene Regulatory Proteins⁵

How does a cell determine which of its thousands of genes to transcribe? As discussed in Chapter 8, the transcription of each gene is controlled by a regulatory region of DNA near the site where transcription begins. Some regulatory regions are simple and act as switches that are thrown by a single signal. Other regulatory regions are complex and act as tiny microprocessors, responding to a variety of signals that they interpret and integrate to switch the neighboring gene on or off. Whether complex or simple, these switching devices consist of two fundamental types of components: (1) short stretches of DNA of defined sequence and (2) *gene regulatory proteins* that recognize and bind to them.

We begin our discussion of gene regulatory proteins by describing how these proteins were discovered.

Gene Regulatory Proteins Were Discovered Using Bacterial Genetics⁶

Genetic analyses in bacteria carried out in the 1950s provided the first evidence of the existence of **gene regulatory proteins** that turn specific sets of genes on or off. One of these regulators, the *lambda repressor*, is encoded by a bacterial virus, *bacteriophage lambda*. The repressor shuts off the viral genes that code for the protein components of new virus particles and thereby enables the viral genome to remain a silent passenger in the bacterial chromosome, multiplying with the bacterium when conditions are favorable for bacterial growth (see Figure 6-80). The lambda repressor was among the first gene regulatory proteins to be characterized, and it remains one of the best understood, as we discuss later. Other bacterial regulators respond to nutritional conditions by shutting off genes encoding specific sets of metabolic enzymes when they are not needed. The *lac repressor*, for example, the first of these bacterial proteins to be recognized, turns off the production of the proteins responsible for lactose metabolism when this sugar is absent from the medium.

The first step toward understanding gene regulation was the isolation of mutant strains of bacteria and bacteriophage lambda that were unable to shut off specific sets of genes. It was proposed at the time, and later proved, that most of these mutants were deficient in proteins acting as specific repressors for these sets of genes. Because these proteins, like most gene regulatory proteins, are present in small quantities, it was difficult and time-consuming to isolate them. They were eventually purified by fractionating cell extracts on a series of standard chromatography columns (see pp. 166-169). Once isolated, the proteins were shown to bind to specific DNA sequences close to the genes that they

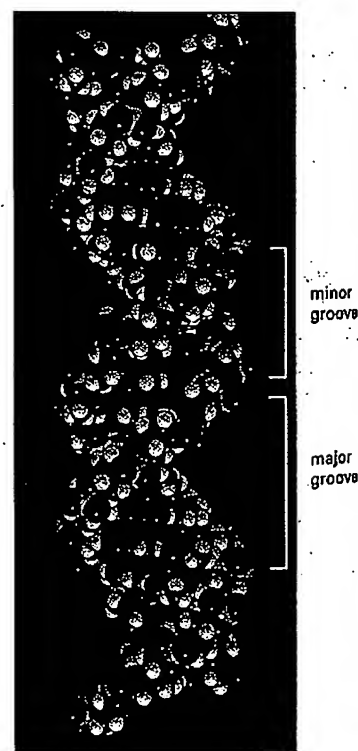
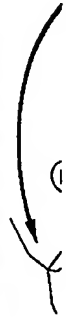
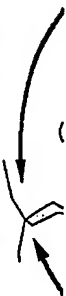


Figure 9-3 Double-helical structure of DNA. The major and minor grooves on the outside of the double helix are indicated. The atoms are colored as follows: carbon, dark blue; nitrogen, light blue; hydrogen, white; oxygen, red; phosphorus, yellow.

regulate
by a cor
experim

The O

As disci
double l
otide se
these pi
pairs in
and anc
ded wit
without
at the s
bond d
to reco:
major g
(Figure
for gro
Alt
most ir
only or
double



DNA.

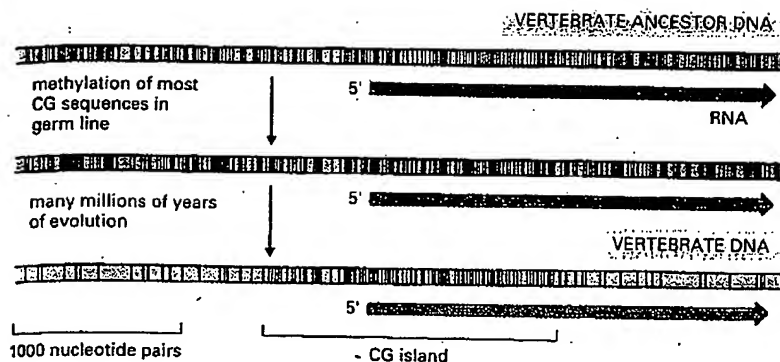


Figure 9-71 A mechanism to explain both the marked deficiency of CG sequences and the presence of CG islands in vertebrate genomes. A black line marks the location of an unmethylated CG dinucleotide in the DNA sequence, while a red line marks the location of a methylated CG dinucleotide.

Summary

The many types of cells in animals and plants are created largely through mechanisms that cause different genes to be transcribed in different cells. Since many specialized animal cells can maintain their unique character when grown in culture, the gene regulatory mechanisms involved in creating them must be stable once established and heritable when the cell divides, endowing the cell with a memory of its developmental history. Prokaryotes and yeasts provide unusually accessible model systems in which to study gene regulatory mechanisms, some of which may be relevant to the creation of specialized cell types in higher eucaryotes. One such mechanism involves a competitive interaction between two (or more) gene regulatory proteins, each of which inhibits the synthesis of the other; this can create a flip-flop switch that switches a cell between two alternative patterns of gene expression. Direct or indirect positive feedback loops, which enable gene regulatory proteins to perpetuate their own synthesis, provide a general mechanism for cell memory.

In eucaryotes gene transcription is generally controlled by combinations of gene regulatory proteins. It is thought that each type of cell in a higher eucaryotic organism contains a specific combination of gene regulatory proteins that ensures the expression of only those genes appropriate to that type of cell. A given gene regulatory protein may be expressed in a variety of circumstances and typically is involved in the regulation of many genes.

In addition to diffusible gene regulatory proteins, inherited states of chromatin condensation are also utilized by eucaryotic cells to regulate gene expression. In vertebrates DNA methylation also plays a part, mainly as a device to reinforce decisions about gene expression that are made initially by other mechanisms.

Posttranscriptional Controls

Although controls on the initiation of gene transcription are the predominant form of regulation for most genes, other controls can act later in the pathway from RNA to protein to modulate the amount of gene product that is made. Although these posttranscriptional controls, which operate after RNA polymerase has bound to the gene's promoter and begun RNA synthesis, are less common than transcriptional control, for many genes they are crucial. It seems that every step in gene expression that could be controlled in principle is likely to be regulated under some circumstances for some genes.

We consider the varieties of posttranscriptional regulation in temporal order, according to the sequence of events that might be experienced by an RNA molecule after its transcription has begun (Figure 9-72).

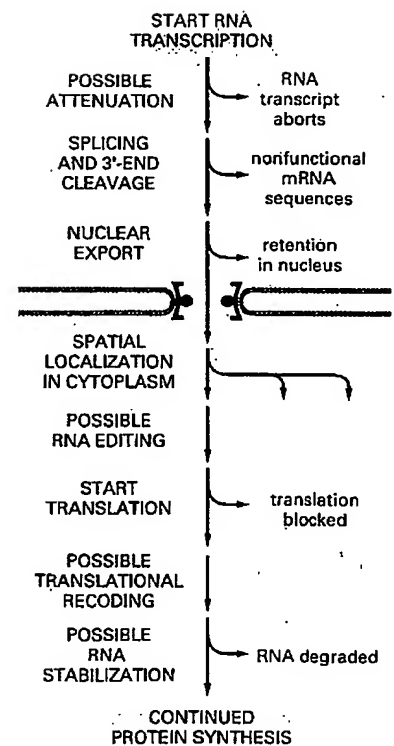


Figure 9-72 Possible posttranscriptional controls on gene expression. Only a few of these controls are likely to be used for any one gene.



MOLECULAR BIOLOGY OF
THE CELL

fourth edition

Bruce Alberts

Alexander Johnson

Julian Lewis

Martin Raff

Keith Roberts

Peter Walter

GS Garland Science
Taylor & Francis Group

Garland

Vice President: Denise Schanck
Managing Editor: Sarah Gibbs
Senior Editorial Assistant: Kirsten Jenner
Managing Production Editor: Emma Hunt
Proofreader and Layout: Emma Hunt
Production Assistant: Angela Bennett
Text Editors: Marjorie Singer Anderson and Betsy Dilemnia
Copy Editor: Bruce Goatly
Word Processors: Fran Dependahl, Misty Landers and Carol Winter
Designer: Blink Studio, London
Illustrator: Nigel Orme
Indexer: Janine Ross and Sherry Granum
Manufacturing: Nigel Eyre and Marion Morrow

Bruce Alberts received his Ph.D. from Harvard University and is President of the National Academy of Sciences and Professor of Biochemistry and Biophysics at the University of California, San Francisco. **Alexander Johnson** received his Ph.D. from Harvard University and is a Professor of Microbiology and Immunology at the University of California, San Francisco. **Julian Lewis** received his D.Phil. from the University of Oxford and is a Principal Scientist at the Imperial Cancer Research Fund, London. **Martin Raff** received his M.D. from McGill University and is at the Medical Research Council Laboratory for Molecular Cell Biology and Cell Biology Unit and in the Biology Department at University College London. **Keith Roberts** received his Ph.D. from the University of Cambridge and is Associate Research Director at the John Innes Centre, Norwich. **Peter Walter** received his Ph.D. from The Rockefeller University in New York and is Professor and Chairman of the Department of Biochemistry and Biophysics at the University of California, San Francisco, and an Investigator of the Howard Hughes Medical Institute.

© 2002 by Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter.
© 1983, 1989, 1994 by Bruce Alberts, Dennis Bray, Julian Lewis, Martin Raff, Keith Roberts, and James D. Watson.

All rights reserved. No part of this book covered by the copyright hereon may be reproduced or used in any format in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—without permission of the publisher.

Library of Congress Cataloging-in-Publication Data

Molecular biology of the cell / Bruce Alberts ... [et al.]. -- 4th ed.
p. cm
Includes bibliographical references and index.
ISBN 0-8153-3218-1 (hardbound) -- ISBN 0-8153-4072-9 (pbk.)
1. Cytology. 2. Molecular biology. I. Alberts, Bruce.
[DNLM: 1. Cells. 2. Molecular Biology.]
QH581.2 .M64 2002
571.6--dc21

2001054471 CIP

Published by Garland Science, a member of the Taylor & Francis Group,
29 West 35th Street, New York, NY 10001-2299

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

Cell Biology Interactive

Artistic and Scientific Direction: Peter Walter
Narrated by: Julie Theriot
Production, Design, and Development: Mike Morales

Front cover Human Genome: Reprinted by permission from *Nature*, International Human Genome Sequencing Consortium, 409:860–921, 2001 © Macmillan Magazines Ltd. Adapted from an image by Francis Collins, NHGRI; Jim Kent, UCSC; Ewan Birney, EBI; and Darryl Leja, NHGRI; showing a portion of Chromosome 1 from the initial sequencing of the human genome.

Back cover In 1967, the British artist Peter Blake created a design classic. Nearly 35 years later Nigel Orme (illustrator), Richard Denyer (photographer), and the authors have together produced an affectionate tribute to Mr Blake's image. With its gallery of icons and influences, its assembly created almost as much complexity, intrigue and mystery as the original. *Drosophila*, *Arabidopsis*, Dolly and the assembled company tempt you to dip inside where, as in the original, "a splendid time is guaranteed for all." (Gunter Blobel, courtesy of The Rockefeller University; Marie Curie, Keystone Press Agency Inc; Darwin bust, by permission of the President and Council of the Royal Society; Rosalind Franklin, courtesy of Cold Spring Harbor Laboratory Archives; Dorothy Hodgkin, © The Nobel Foundation, 1984; James Joyce, etching by Peter Blake; Robert Johnson, photo booth self-portrait early 1930s, © 1986 Delta Haze Corporation all rights reserved, used by permission; Albert L. Lehninger, (unidentified photographer) courtesy of The Alan Mason Chesney Medical Archives of The Johns Hopkins Medical Institutions; Linus Pauling, from Ava Helen and Linus Pauling Papers, Special Collections, Oregon State University; Nicholas Poussin, courtesy of ArtToday.com; Barbara McClintock, © David Micklos, 1983; Andrei Sakharov, courtesy of Elena Bonner; Frederick Sanger, © The Nobel Foundation, 1958.)

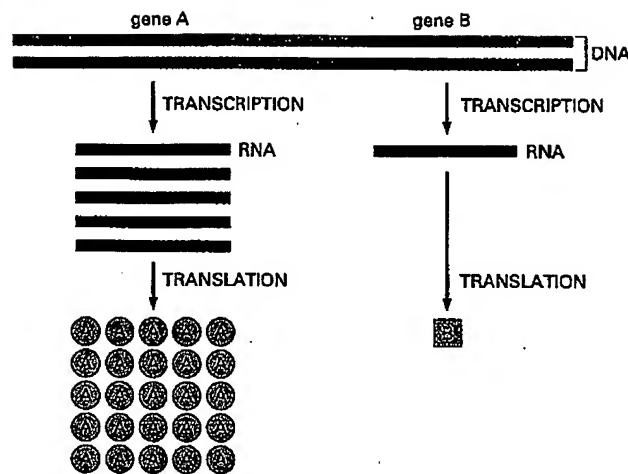


Figure 6-3 Genes can be expressed with different efficiencies. Gene A is transcribed and translated much more efficiently than gene B. This allows the amount of protein A in the cell to be much greater than that of protein B.

FROM DNA TO RNA

Transcription and translation are the means by which cells read out, or express, the genetic instructions in their genes. Because many identical RNA copies can be made from the same gene, and each RNA molecule can direct the synthesis of many identical protein molecules, cells can synthesize a large amount of protein rapidly when necessary. But each gene can also be transcribed and translated with a different efficiency, allowing the cell to make vast quantities of some proteins and tiny quantities of others (Figure 6-3). Moreover, as we see in the next chapter, a cell can change (or regulate) the expression of each of its genes according to the needs of the moment—most obviously by controlling the production of its RNA.

Portions of DNA Sequence Are Transcribed into RNA

The first step a cell takes in reading out a needed part of its genetic instructions is to copy a particular portion of its DNA nucleotide sequence—a gene—into an RNA nucleotide sequence. The information in RNA, although copied into another chemical form, is still written in essentially the same language as it is in DNA—the language of a nucleotide sequence. Hence the name **transcription**.

Like DNA, RNA is a linear polymer made of four different types of nucleotide subunits linked together by phosphodiester bonds (Figure 6-4). It differs from DNA chemically in two respects: (1) the nucleotides in RNA are *ribonucleotides*—that is, they contain the sugar ribose (hence the name *ribonucleic acid*) rather than deoxyribose; (2) although, like DNA, RNA contains the bases adenine (A), guanine (G), and cytosine (C), it contains the base uracil (U) instead of the thymine (T) in DNA. Since U, like T, can base-pair by hydrogen-bonding with A (Figure 6-5), the complementary base-pairing properties described for DNA in Chapters 4 and 5 apply also to RNA (in RNA, G pairs with C, and A pairs with U). It is not uncommon, however, to find other types of base pairs in RNA: for example, G pairing with U occasionally.

Despite these small chemical differences, DNA and RNA differ quite dramatically in overall structure. Whereas DNA always occurs in cells as a double-stranded helix, RNA is single-stranded. RNA chains therefore fold up into a variety of shapes, just as a polypeptide chain folds up to form the final shape of a protein (Figure 6-6). As we see later in this chapter, the ability to fold into complex three-dimensional shapes allows some RNA molecules to have structural and catalytic functions.

Transcription Produces RNA Complementary to One Strand of DNA

All of the RNA in a cell is made by DNA transcription, a process that has certain similarities to the process of DNA replication discussed in Chapter 5.

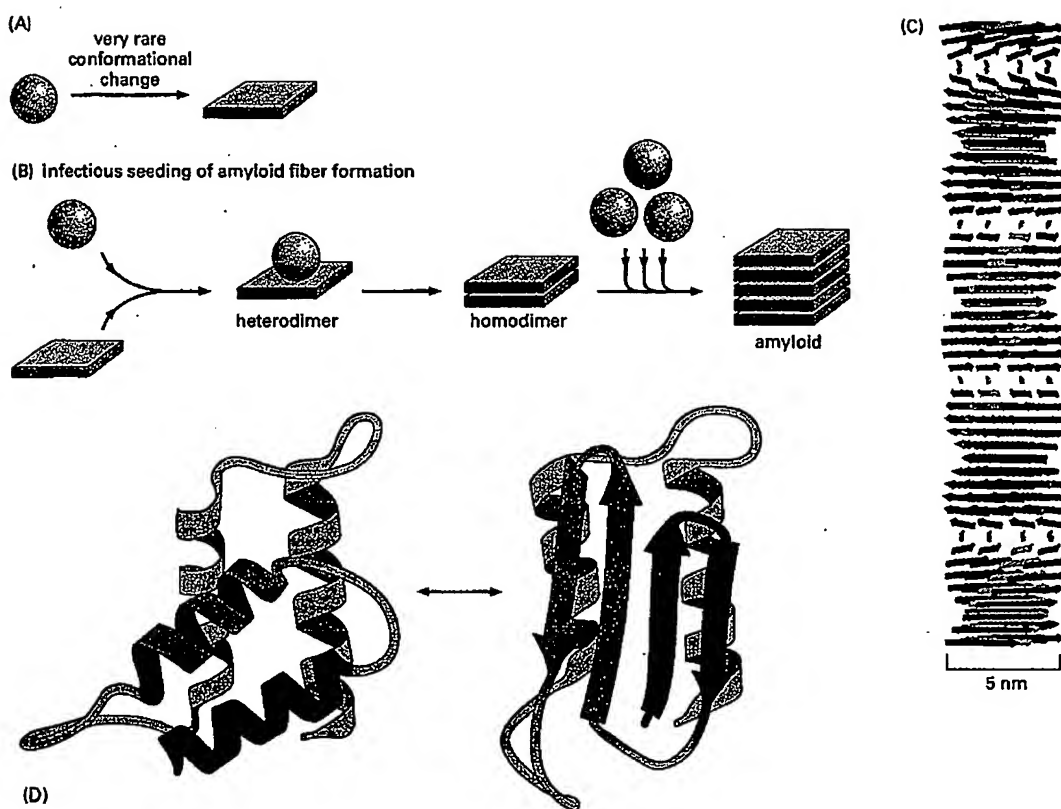


Figure 6-89 Protein aggregates that cause human disease. (A) Schematic illustration of the type of conformational change in a protein that produces material for a cross-beta filament. (B) Diagram illustrating the self-infectious nature of the protein aggregation that is central to prion diseases. PrP is highly unusual because the misfolded version of the protein, called PrP^{*}, induces the normal PrP protein it contacts to change its conformation, as shown. Most of the human diseases caused by protein aggregation are caused by the overproduction of a variant protein that is especially prone to aggregation, but because this structure is not infectious in this way, it cannot spread from one animal to another. (C) Drawing of a cross-beta filament, a common type of protease-resistant protein aggregate found in a variety of human neurological diseases. Because the hydrogen-bond interactions in a β sheet form between polypeptide backbone atoms (see Figure 3-9), a number of different abnormally folded proteins can produce this structure. (D) One of several possible models for the conversion of PrP to PrP^{*}, showing the likely change of two α -helices into four β -strands. Although the structure of the normal protein has been determined accurately, the structure of the infectious form is not yet known with certainty because the aggregation has prevented the use of standard structural techniques. (C, courtesy of Louise Serpell, adapted from M. Sunde et al., *J. Mol. Biol.* 273:729-739, 1997; D, adapted from S.B. Prusiner, *Trends Biochem. Sci.* 21:482-487, 1996.)

animals and humans. It can be dangerous to eat the tissues of animals that contain PrP^{*}, as witnessed most recently by the spread of BSE (commonly referred to as the "mad cow disease") from cattle to humans in Great Britain.

Fortunately, in the absence of PrP^{*}, PrP is extraordinarily difficult to convert to its abnormal form. Although very few proteins have the potential to misfold into an infectious conformation, a similar transformation has been discovered to be the cause of an otherwise mysterious "protein-only inheritance" observed in yeast cells.

There Are Many Steps From DNA to Protein

We have seen so far in this chapter that many different types of chemical reactions are required to produce a properly folded protein from the information contained in a gene (Figure 6-90). The final level of a properly folded protein in a cell therefore depends upon the efficiency with which each of the many steps is performed.

We discuss in Chapter 7 that cells have the ability to change the levels of their proteins according to their needs. In principle, any or all of the steps in Fig-

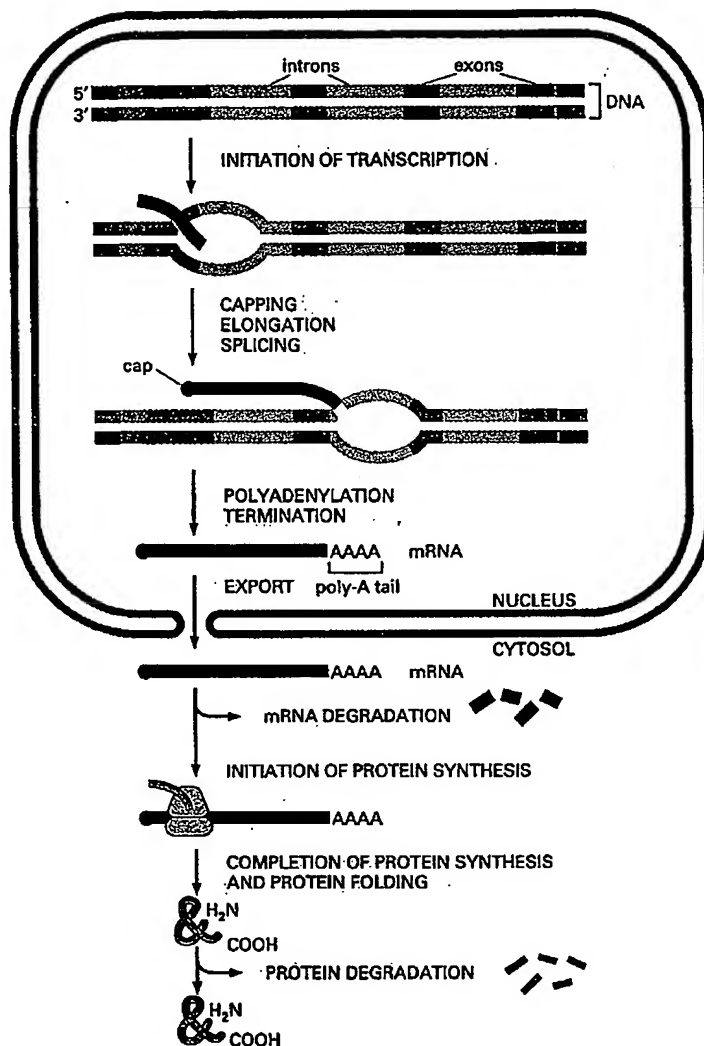


Figure 6-90 The production of a protein by a eucaryotic cell. The final level of each protein in a eucaryotic cell depends upon the efficiency of each step depicted.

ure 6-90) could be regulated by the cell for each individual protein. However, as we shall see in Chapter 7, the initiation of transcription is the most common point for a cell to regulate the expression of each of its genes. This makes sense, inasmuch as the most efficient way to keep a gene from being expressed is to block the very first step—the transcription of its DNA sequence into an RNA molecule.

Summary

The translation of the nucleotide sequence of an mRNA molecule into protein takes place in the cytoplasm on a large ribonucleoprotein assembly called a ribosome. The amino acids used for protein synthesis are first attached to a family of tRNA molecules, each of which recognizes, by complementary base-pair interactions, particular sets of three nucleotides in the mRNA (codons). The sequence of nucleotides in the mRNA is then read from one end to the other in sets of three according to the genetic code.

To initiate translation, a small ribosomal subunit binds to the mRNA molecule at a start codon (AUG) that is recognized by a unique initiator tRNA molecule. A large ribosomal subunit binds to complete the ribosome and begin the elongation phase of protein synthesis. During this phase, aminoacyl tRNAs—each bearing a specific amino acid bind sequentially to the appropriate codon in mRNA by forming complementary base pairs with the tRNA anticodon. Each amino acid is added to the C-terminal end of the growing polypeptide by means of a cycle of three sequential

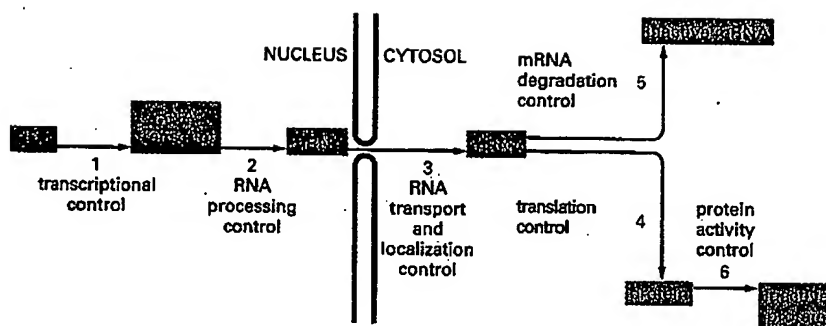


Figure 7-5 Six steps at which eucaryotic gene expression can be controlled. Controls that operate at steps 1 through 5 are discussed in this chapter. Step 6, the regulation of protein activity, includes reversible activation or inactivation by protein phosphorylation (discussed in Chapter 3) as well as irreversible inactivation by proteolytic degradation (discussed in Chapter 6).

Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein

If differences among the various cell types of an organism depend on the particular genes that the cells express, at what level is the control of gene expression exercised? As we saw in the last chapter, there are many steps in the pathway leading from DNA to protein, and all of them can in principle be regulated. Thus a cell can control the proteins it makes by (1) controlling when and how often a given gene is transcribed (**transcriptional control**), (2) controlling how the RNA transcript is spliced or otherwise processed (**RNA processing control**), (3) selecting which completed mRNAs in the cell nucleus are exported to the cytosol and determining where in the cytosol they are localized (**RNA transport and localization control**), (4) selecting which mRNAs in the cytoplasm are translated by ribosomes (**translational control**), (5) selectively destabilizing certain mRNA molecules in the cytoplasm (**mRNA degradation control**), or (6) selectively activating, inactivating, degrading, or compartmentalizing specific protein molecules after they have been made (**protein activity control**) (Figure 7-5).

For most genes transcriptional controls are paramount. This makes sense because, of all the possible control points illustrated in Figure 7-5, only transcriptional control ensures that the cell will not synthesize superfluous intermediates. In the following sections we discuss the DNA and protein components that perform this function by regulating the initiation of gene transcription. We shall return at the end of the chapter to the additional ways of regulating gene expression.

Summary

The genome of a cell contains in its DNA sequence the information to make many thousands of different protein and RNA molecules. A cell typically expresses only a fraction of its genes, and the different types of cells in multicellular organisms arise because different sets of genes are expressed. Moreover, cells can change the pattern of genes they express in response to changes in their environment, such as signals from other cells. Although all of the steps involved in expressing a gene can in principle be regulated, for most genes the initiation of RNA transcription is the most important point of control.

DNA-BINDING MOTIFS IN GENE REGULATORY PROTEINS

How does a cell determine which of its thousands of genes to transcribe? As mentioned briefly in Chapters 4 and 6, the transcription of each gene is controlled by a regulatory region of DNA relatively near the site where transcription begins. Some regulatory regions are simple and act as switches that are thrown by a single signal. Many others are complex and act as tiny microprocessors, responding to a variety of signals that they interpret and integrate to switch the neighboring gene on or off. Whether complex or simple, these switching devices

occur in the germ line, the cell lineage that gives rise to sperm or eggs. Most of the DNA in vertebrate germ cells is inactive and highly methylated. Over long periods of evolutionary time, the methylated CG sequences in these inactive regions have presumably been lost through spontaneous deamination events that were not properly repaired. However promoters of genes that remain active in the germ cell lineages (including most housekeeping genes) are kept unmethylated, and therefore spontaneous deaminations of Cs that occur within them can be accurately repaired. Such regions are preserved in modern day vertebrate cells as CG islands. In addition, any mutation of a CG sequence in the genome that destroyed the function or regulation of a gene in the adult would be selected against, and some CG islands are simply the result of a higher than normal density of critical CG sequences.

The mammalian genome contains an estimated 20,000 CG islands. Most of the islands mark the 5' ends of transcription units and thus, presumably, of genes. The presence of CG islands often provides a convenient way of identifying genes in the DNA sequences of vertebrate genomes.

Summary

The many types of cells in animals and plants are created largely through mechanisms that cause different genes to be transcribed in different cells. Since many specialized animal cells can maintain their unique character through many cell division cycles and even when grown in culture, the gene regulatory mechanisms involved in creating them must be stable once established and heritable when the cell divides. These features endow the cell with a memory of its developmental history. Bacteria and yeasts provide unusually accessible model systems in which to study gene regulatory mechanisms. One such mechanism involves a competitive interaction between two gene regulatory proteins, each of which inhibits the synthesis of the other; this can create a flip-flop switch that switches a cell between two alternative patterns of gene expression. Direct or indirect positive feedback loops, which enable gene regulatory proteins to perpetuate their own synthesis, provide a general mechanism for cell memory. Negative feedback loops with programmed delays form the basis for cellular clocks.

In eucaryotes the transcription of a gene is generally controlled by combinations of gene regulatory proteins. It is thought that each type of cell in a higher eucaryotic organism contains a specific combination of gene regulatory proteins that ensures the expression of only those genes appropriate to that type of cell. A given gene regulatory protein may be active in a variety of circumstances and typically is involved in the regulation of many genes.

In addition to diffusible gene regulatory proteins, inherited states of chromatin condensation are also used by eucaryotic cells to regulate gene expression. An especially dramatic case is the inactivation of an entire X chromosome in female mammals. In vertebrates DNA methylation also functions in gene regulation, being used mainly as a device to reinforce decisions about gene expression that are made initially by other mechanisms. DNA methylation also underlies the phenomenon of genomic imprinting in mammals, in which the expression of a gene depends on whether it was inherited from the mother or the father.

POSTTRANSCRIPTIONAL CONTROLS

In principle, every step required for the process of gene expression could be controlled. Indeed, one can find examples of each type of regulation, although any one gene is likely to use only a few of them. Controls on the initiation of gene transcription are the predominant form of regulation for most genes. But other controls can act later in the pathway from DNA to protein to modulate the amount of gene product that is made. Although these **posttranscriptional controls**, which operate after RNA polymerase has bound to the gene's promoter and begun RNA synthesis, are less common than *transcriptional control*, for many genes they are crucial.

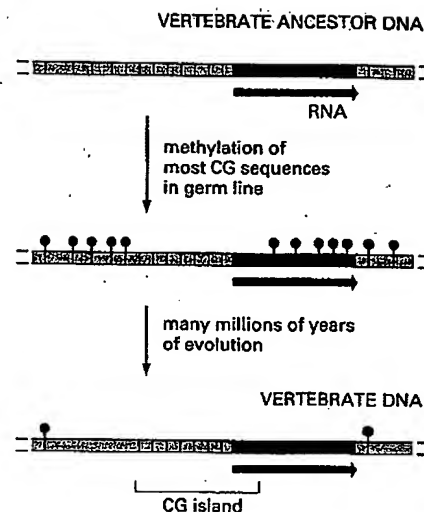


Figure 7-86 A mechanism to explain both the marked overall deficiency of CG sequences and their clustering into CG islands in vertebrate genomes. A black line marks the location of a CG dinucleotide in the DNA sequence, while a red "lollipop" indicates the presence of a methyl group on the CG dinucleotide. CG sequences that lie in regulatory sequences of genes that are transcribed in germ cells are unmethylated and therefore tend to be retained in evolution. Methylated CG sequences, on the other hand, tend to be lost through deamination of 5-methyl C to T, unless the CG sequence is critical for survival.

CHAPTER 29

Regulation of transcription

Genes VI (1997) CH 29, pp. 847-848.
Benjamin Lewin

The phenotypic differences that distinguish the various kinds of cells in a higher eukaryote are largely due to differences in the expression of genes that code for proteins, that is, those transcribed by RNA polymerase II. In principle, the expression of these genes might be regulated at any one of several stages. The concept of the "level of control" implies that gene expression is not necessarily an automatic process once it has begun. It could be regulated in a gene-specific way at any one of several sequential steps. We can distinguish (at least) five potential control points, forming the series:

Activation of gene structure
↓
Initiation of transcription
↓
Processing the transcript
↓
Transport to cytoplasm
↓
Translation of mRNA

The existence of the first step is implied by the discovery that genes may exist in either of two structural conditions. Relative to the state of most of the genome, genes are found in an "active" state in the cells in which they are expressed (see Chapter 27). The change of structure is distinct from the act of transcription, and indicates that the gene is "transcribable." This suggests that acquisition of the "active" structure must be the first step in gene expression.

Transcription of a gene in the active state is

controlled at the stage of initiation, that is, by the interaction of RNA polymerase with its promoter. This is now becoming susceptible to analysis in the *in vitro* systems (see Chapter 28). For most genes, this is a major control point; probably it is the most common level of regulation.

There is at present no evidence for control at subsequent stages of transcription in eukaryotic cells, for example, via antitermination mechanisms.

The primary transcript is modified by capping at the 5' end, and usually also by polyadenylation at the 3' end. Introns must be spliced out from the transcripts of interrupted genes. The mature RNA must be exported from the nucleus to the cytoplasm. Regulation of gene expression by selection of sequences at the level of nuclear RNA might involve any or all of these stages, but the one for which we have most evidence concerns changes in splicing; some genes are expressed by means of alternative splicing patterns whose regulation controls the type of protein product (see Chapter 30).

Finally, the translation of an mRNA in the cytoplasm can be specifically controlled. There is little evidence for the employment of this mechanism in adult somatic cells, but it does occur in some embryonic situations, as described in Chapter 7. The mechanism is presumed to involve the blocking of initiation of translation of some mRNAs by specific protein factors.

But having acknowledged that control of gene expression can occur at multiple stages, and that production of RNA cannot inevitably be equated with production of protein, it is clear

that the overwhelming majority of regulatory events occur at the initiation of transcription. Regulation of tissue-specific gene transcription lies at the heart of eukaryotic differentiation; indeed, we see examples in Chapter 38 in which proteins that regulate embryonic development prove to be transcription factors. A regulatory transcription factor serves to provide

common control of a large number of target genes, and we seek to answer two questions about this mode of regulation: what identifies the common target genes to the transcription factor; and how is the activity of the transcription factor itself regulated in response to intrinsic or extrinsic signals?

Response elements identify genes under common regulation

The principle that emerges from characterizing groups of genes under common control is that *they share a promoter element that is recognized by a regulatory transcription factor*. An element that causes a gene to respond to such a factor is called a response element; examples are the HSE (heat shock response element), GRE (glucocorticoid response element), SRE (serum response element).

The properties of some inducible transcription factors and the elements that they recognize are summarized in Table 29.1. Response elements have the same general characteristics as upstream elements of promoters or enhancers. They contain short consensus sequences, and copies of the response elements found in different genes are closely related, but not necessarily identical. The region bound by the factor extends for a short distance on either side of

the consensus sequence. In promoters, the elements are not present at fixed distances from the startpoint, but are usually <200 bp upstream of it. The presence of a single element usually is sufficient to confer the regulatory response, but sometimes there are multiple copies.

Response elements may be located in promoters or in enhancers. Some types of elements are typically found in one rather than the other: usually an HSE is found in a promoter, while a GRE is found in an enhancer. We assume that all response elements function by the same general principle. *A gene is regulated by a sequence at the promoter or enhancer that is recognized by a specific protein. The protein functions as a transcription factor needed for RNA polymerase to initiate. Active protein is available only under conditions when the gene is to be expressed; its absence means that the promoter is not activated by this particular circuit.*

An example of a situation in which many genes are controlled by a single factor is provided by the heat shock response. This is common to a wide range of prokaryotes and eukaryotes and involves multiple controls of gene expression: an increase in temperature turns off transcription of some genes, turns on transcription of the heat shock genes, and causes changes in the translation of mRNAs. The control of the heat shock genes illustrates the differences between prokaryotic and eukaryotic modes of control. In bacteria, a new sigma factor is synthesized that directs RNA polymerase holoenzyme to recognize an *alter-*

Table 29.1 Inducible transcription factors bind to response elements that identify groups of promoters or enhancers subject to coordinate control.

Regulatory Agent	Module	Consensus	Factor
Heat shock	HSE	CNNGAANNTCGNG	HSTF
Glucocorticoid	GRE	TGGTACAAATGTTCT	Receptor
Phorbol ester	TRE	TGACTCA	AP1
Serum	SRE	CCATATTAGG	SRF

Research

Open Access

Prostate stem cell antigen (PSCA) expression in human prostate cancer tissues and its potential role in prostate carcinogenesis and progression of prostate cancer

Zhao Zhigang*¹ and Shen Wenlv²

Address: ¹Department of Urology, Shantou University Medical College, Shantou, Guangdong, China and ²Department of Urology, No 2. Affiliated Hospital of Shantou University Medical College, Shantou, Guangdong, China

Email: Zhao Zhigang* - zgzhao@yeah.com.cn; Shen Wenlv - wshen99@hotmail.com

* Corresponding author

Published: 10 May 2004

Received: 30 March 2004

World Journal of Surgical Oncology 2004, 2:13

Accepted: 10 May 2004

This article is available from: <http://www.wjso.com/content/2/1/13>

© 2004 Zhigang and Wenlv; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Prostate stem cell antigen (PSCA) is a recently defined homologue of the Thy-1/Ly-6 family of glycosylphosphatidylinositol (GPI)-anchored cell surface antigens. The purpose of the present study was to examine the expression status of PSCA protein and mRNA in clinical specimens of human prostate cancer (Pca) and to validate it as a potential molecular target for diagnosis and treatment of Pca.

Materials and Methods: Immunohistochemical (IHC) and *in situ* hybridization (ISH) analyses of PSCA expression were simultaneously performed on paraffin-embedded sections from 20 benign prostatic hyperplasia (BPH), 20 prostatic intraepithelial neoplasm (PIN) and 48 prostate cancer (Pca) tissues, including 9 androgen-independent prostate cancers. The level of PSCA expression was semiquantitatively scored by assessing both the percentage and intensity of PSCA-positive staining cells in the specimens. Then compared PSCA expression between BPH, PIN and Pca tissues and analysed the correlations of PSCA expression level with pathological grade, clinical stage and progression to androgen-independence in Pca.

Results: In BPH and low grade PIN, PSCA protein and mRNA staining were weak or negative and less intense and uniform than that seen in HGPIN and Pca. There were moderate to strong PSCA protein and mRNA expression in 8 of 11 (72.7%) HGPIN and in 40 of 48 (83.4%) Pca specimens examined by IHC and ISH analyses, with statistical significance compared with BPH (20%) and low grade PIN (22.2%) samples ($p < 0.05$, respectively). The expression level of PSCA increased with high Gleason grade, advanced stage and progression to androgen-independence ($p < 0.05$, respectively). In addition, IHC and ISH staining showed a high degree of correlation between PSCA protein and mRNA overexpression.

Conclusions: Our data demonstrate that PSCA as a new cell surface marker is overexpressed by a majority of human Pca. PSCA expression correlates positively with adverse tumor characteristics, such as increasing pathological grade (poor cell differentiation), worsening clinical stage and androgen-independence, and speculatively with prostate carcinogenesis. PSCA protein overexpression results from upregulated transcription of PSCA mRNA. PSCA may have prognostic utility and may be a promising molecular target for diagnosis and treatment of Pca.

Introduction

Prostate cancer (Pca) is the second leading cause of cancer-related death in American men and is becoming a common cancer increasing in China. Despite recently great progress in the diagnosis and management of localized disease, there continues to be a need for new diagnostic markers that can accurately discriminate between indolent and aggressive variants of Pca. There also continues to be a need for the identification and characterization of potential new therapeutic targets on Pca cells. Current diagnostic and therapeutic modalities for recurrent and metastatic Pca have been limited by a lack of specific target antigens of Pca.

Although a number of prostate-specific genes have been identified (i.e. prostate specific antigen, prostatic acid phosphatase, glandular kallikrein 2), the majority of these are secreted proteins not ideally suited for many immunological strategies. So, the identification of new cell surface antigens is critical to the development of new diagnostic and therapeutic approaches to the management of Pca.

Reiter RE et al [1] reported the identification of prostate stem cell antigen (PSCA), a cell surface antigen that is predominantly prostate specific. The PSCA gene encodes a 123 amino acid glycoprotein, with 30% homology to stem cell antigen 2 (Sca 2). Like Sca-2, PSCA also belongs to a member of the Thy-1/Ly-6 family and is anchored by a glycosylphosphatidylinositol (GPI) linkage. mRNA *in situ* hybridization (ISH) localized PSCA expression in normal prostate to the basal cell epithelium, the putative stem cell compartment of prostatic epithelium, suggesting that PSCA may be a marker of prostate stem/progenitor cells.

In order to examine the status of PSCA protein and mRNA expression in human Pca and validate it as a potential diagnostic and therapeutic target for Pca, we used immunohistochemistry (IHC) and *in situ* hybridization (ISH) simultaneously, and conducted PSCA protein and mRNA expression analyses in paraffin-embedded tissue specimens of benign prostatic hyperplasia (BPH, n = 20), prostate intraepithelial neoplasm (PIN, n = 20) and prostate cancer (Pca, n = 48). Furthermore, we evaluated the possible correlation of PSCA expression level with Pca tumorigenesis, grade, stage and progression to androgen-independence.

Materials and methods

Tissue samples

All of the clinical tissue specimens studied herein were obtained from 80 patients of 57-84 years old by prostatectomy, transurethral resection of prostate (TURP) or biopsies. The patients were classified as 20 cases of BPH, 20 cases of PIN, 40 cases of primary Pca, including 9 patients

with recurrent Pca and a history of androgen ablation therapy (orchiectomy and/or hormonal therapy), who were referred to as androgen-independent prostate cancers. Eight specimens were harvested from these androgen-independent Pca patients prior to androgen ablation treatment. Each tissue sample was cut into two parts, one was fixed in 10% formalin for IHC and the other treated with 4% paraformaldehyde/0.1 M PBS PH 7.4 in 0.1% DEPC for 1 h for ISH analysis, and then embedded in paraffin. All paraffin blocks examined were then cut into 5 μ m sections and mounted on the glass slides specific for IHC and ISH respectively in the usual fashion. H&E-stained section of each Pca was evaluated and assigned a Gleason score by the experienced urological pathologist at our institution based on the criteria of Gleason score [2]. The Gleason sums are summarized in Table 1. Clinical staging was performed according to Jewett-whitmore-prout staging system, as shown in Table 2. In the category of PIN, we graded the specimens into two groups, i.e. low grade PIN (grade I - II) and high grade PIN (HGPN, grade III) on the basis of literatures [3,4].

Immunohistochemical (IHC) analysis

Briefly, tissue sections were deparaffinized, dehydrated, and subjected to microwaving in 10 mmol/L citrate buffer, PH 6.0 (Boshide, Wuhan, China) in a 900 W oven for 5 min to induce epitope retrieval. Slides were allowed to cool at room temperature for 30 min. A primary mouse antibody specific to human PSCA (Boshide, Wuhan, China) with a 1:100 dilution was applied to incubate with the slides at room temperature for 2 h. Labeling was detected by sequentially adding biotinylated secondary antibodies and streptavidin-peroxidase, and localized using 3,3'-diaminobenzidine reaction. Sections were then counterstained with hematoxylin. Substitution of the primary antibody with phosphate-buffered-saline (PBS) served as a negative-staining control.

mRNA *in situ* hybridization (ISH)

Five- μ m-thick tissue sections were deparaffinized and dehydrated, then digested in pepsin solution (4 mg/ml in 3% citric acid) for 20 min at 37.5°C, and further processed for ISH. Digoxigenin-labeled sense and antisense human PSCA RNA probes (obtained from Boshide, Wuhan, China) were hybridized to the sections at 48°C overnight. The posthybridization wash with a high stringency was performed sequentially at 37°C in 2 \times standard saline citrate (SSC) for 10 min, in 0.5 \times SSC for 15 min and in 0.2 \times SSC for 30 min. The slides were then incubated to biotinylated mouse anti-digoxigenin antibody at 37.5°C for 1 h followed by washing in 1 \times PBS for 20 min at room temperature, and then to streptavidin-peroxidase at 37.5°C for 20 min followed by washing in 1 \times PBS for 15 min at room temperature. Subsequently, the slides were developed with diaminobenzidine and then coun-

Table 1: Correlation of PSCA expression with Gleason score

Gleason score	Intensity × frequency	
	0-6 (%)	9 (%)
2-4	5 (83)	1 (17)
5-7	19 (79)	5 (21)
8-10	5 (28)	13 (72)

Table 2: Correlation of PSCA expression with clinical stage

Tumor stage	Intensity × frequency	
	0-6 (%)	9 (%)
≤B	27 (67.5)	13 (32.5)
≥C	2 (25)	6 (75)

terstained with hematoxylin to localize the hybridization signals. Sections hybridized with the sense control probes routinely did not show any specific hybridization signal above background. All slides were hybridized with PBS to substitute for the probes as a negative control.

Scoring methods

To determine the correlation between the results of PSCA immunostaining and mRNA *in situ* hybridization, the same scoring manners are taken in the present study for PSCA protein staining by IHC and PSCA mRNA staining by ISH. Each slide was read and scored by two independently experienced urological pathologists using Olympus BX-41 light microscopes. The evaluation was done in a blinded fashion. For each section, five areas of similar grade were analyzed semiquantitatively for the fraction of cells staining. Fifty percent of specimens were randomly chosen and rescored to determine the degree of interobserver and intraobserver concordance. There was greater than 95% intra- and interobserver agreement.

The intensity of PSCA expression evaluated microscopically was graded on a scale of 0 to 3+ with 3 being the highest expression observed (0, no staining; 1+, mildly intense; 2+, moderately intense; 3+, severely intense). The staining density was quantified as the percentage of cells staining positive for PSCA with the primary antibody or hybridization probe, as follows: 0 = no staining; 1 = positive staining in <25% of the sample; 2 = positive staining in 25%-50% of the sample; 3 = positive staining in >50%

of the sample. Intensity score (0 to 3+) was multiplied by the density score (0-3) to give an overall score of 0-9 [1,5]. In this way, we were able to differentiate specimens that may have had focal areas of increased staining from those that had diffuse areas of increased staining [6]. The overall score for each specimen was then categorically assigned to one of the following groups: 0 score, negative expression; 1-2 scores, weak expression; 3-6 scores, moderate expression; 9 score, strong expression.

Statistical analysis

Intensity and density of PSCA protein and mRNA expression in BPH, PIN and Pca tissues were compared using the Chi-square and Student's *t*-test. Univariate associations between PSCA expression and Gleason score, clinical stage and progression to androgen-independence were calculated using Fisher's Exact Test. For all analyses, *p* < 0.05 was considered statistically significant.

Results

PSCA expression in BPH

In general, PSCA protein and mRNA were expressed weakly in individual samples of BPH. Some areas of prostate expressed weak levels (composite score 1-2), whereas other areas were completely negative (composite score 0). Four cases (20%) of BPH had moderate expression of PSCA protein and mRNA (composite score 4-6) by IHC and ISH. In 2/20 (10%) BPH specimens, PSCA mRNA expression was moderate (composite score 3-6), but PSCA protein expression was weak (composite score

2) in one and negative (composite score 0) in the other. PSCA expression was localized to the basal and secretory epithelial cells, and prostatic stroma was almost negative staining for PSCA protein and mRNA in all cases examined.

PSCA expression in PIN

In this study, we detected weak or negative expression of PSCA protein and mRNA (≤ 2 scores) in 7 of 9 (77.8%) low grade PIN and in 2 of 11 (18.2%) HGPIN, and moderate expression (3–6 scores) in the rest 2 low grade PIN and 5 of 11 (45.5%) HGPIN. One HGPIN with moderate PSCA mRNA expression (6 score) was found weak staining for PSCA protein (2 score) by IHC. Strong PSCA protein and mRNA expression (9 score) were detected in the remaining 3 of 11 (27.3%) HGPIN. There was a statistically significant difference of PSCA protein and mRNA expression levels observed between HGPIN and BPH ($p < 0.05$), but no statistical difference reached between low grade PIN and BPH ($p > 0.05$).

PSCA expression in Pca

In order to determine if PSCA protein and mRNA can be detected in prostate cancers and if PSCA expression levels are increased in malignant compared with benign glands, Forty-eight paraffin-embedded Pca specimens were analysed by IHC and ISH. It was shown that 19 of 48 (39.6%) Pca samples stained very strongly for PSCA protein and mRNA with a score of 9 and another 21 (43.8%) specimens displayed moderate staining with scores of 4–6 (Figure 1). In addition, 4 specimens with moderate to strong PSCA mRNA expression (scores of 4–9) had weak protein staining (a score of 2) by IHC analyses. Overall, Pca expressed a significantly higher level of PSCA protein and mRNA than any other specimen category in this study ($p < 0.05$, compared with BPH and PIN respectively). The result demonstrates that PSCA protein and mRNA are overexpressed by a majority of human Pca.

Correlation of PSCA expression with Gleason score in Pca

Using the semi-quantitative scoring method as described in Materials and Methods, we compared the expression level of PSCA protein and mRNA with Gleason grade of Pca, as shown in Table 1. Prostate adenocarcinomas were graded by Gleason score as 2–4 scores = well-differentiation, 5–7 scores = moderate-differentiation and 8–10 scores = poor-differentiation [7]. Seventy-two percent of Gleason scores 8–10 prostate cancers had very strong staining of PSCA compared to 21% with Gleason scores 5–7 and 17% with 2–4 respectively, demonstrating that poorly differentiated Pca had significantly stronger expression of PSCA protein and mRNA than moderately and well differentiated tumors ($p < 0.05$). As depicted in Figure 1, IHC and ISH analyses showed that PSCA protein and mRNA expression in several cases of poorly differen-

tiated Pca were particularly prominent, with more intense and uniform staining. The results indicate that PSCA expression increases significantly with higher tumor grade in human Pca.

Correlation of PSCA expression with clinical stage in Pca

With regards to PSCA expression in every stage of Pca, we showed the results in Table 2. Seventy-five percent of locally advanced and node positive cancers (i.e. C-D stages) expressed statistically high levels of PSCA versus 32.5% that were organ confined (i.e. A-B stages) ($p < 0.05$). The data demonstrate that PSCA expression increases significantly with advanced tumor stage in human Pca.

Correlation of PSCA expression with androgen-independent progression of Pca

All 9 specimens of androgen-independent prostate cancers stained positive for PSCA protein and mRNA. Eight specimens were obtained from patients managed prior to androgen ablation therapy. Seven of eight (87.5%) of these androgen-independent prostate cancers were in the strongest staining category (score = 9), compared with three out of eight (37.5%) of patients with androgen-dependent cancers ($p < 0.05$). The results demonstrate that PSCA expression increases significantly with progression to androgen-independence of human Pca.

It is evident from the results above that within a majority of human prostate cancers the level of PSCA protein and mRNA expression correlates significantly with increasing grade, worsening stage and progression to androgen-independence.

Correlation of PSCA immunostaining and mRNA in situ hybridization

In all 88 specimens surveyed herein, we compared the results of PSCA IHC staining with mRNA ISH analysis. Positive staining areas and its intensity and density scores evaluated by IHC were identical to those seen by ISH in 79 of 88 (89.8%) specimens (18/20 BPH, 19/20 PIN and 42/48 Pca respectively). Importantly, 27/27 samples with PSCA mRNA composite scores of 0–2, 32/36 samples with scores of 3–6 and 22/24 samples with a score of 9 also had PSCA protein expression scores of 0–2, 3–6 and 9 respectively. However, in 5 samples with PSCA mRNA overall scores of 3–6 and in 2 with scores of 9 there were less or negative PSCA protein expression (i.e. scores of 0–4), suggesting that this may reflect posttranscriptional modification of PSCA or that the epitopes recognized by PSCA mAb may be obscured in some cancers. The data demonstrate that the results of PSCA immunostaining were consistent with those of mRNA ISH analysis, showing a high degree of correlation between PSCA protein and mRNA expression.

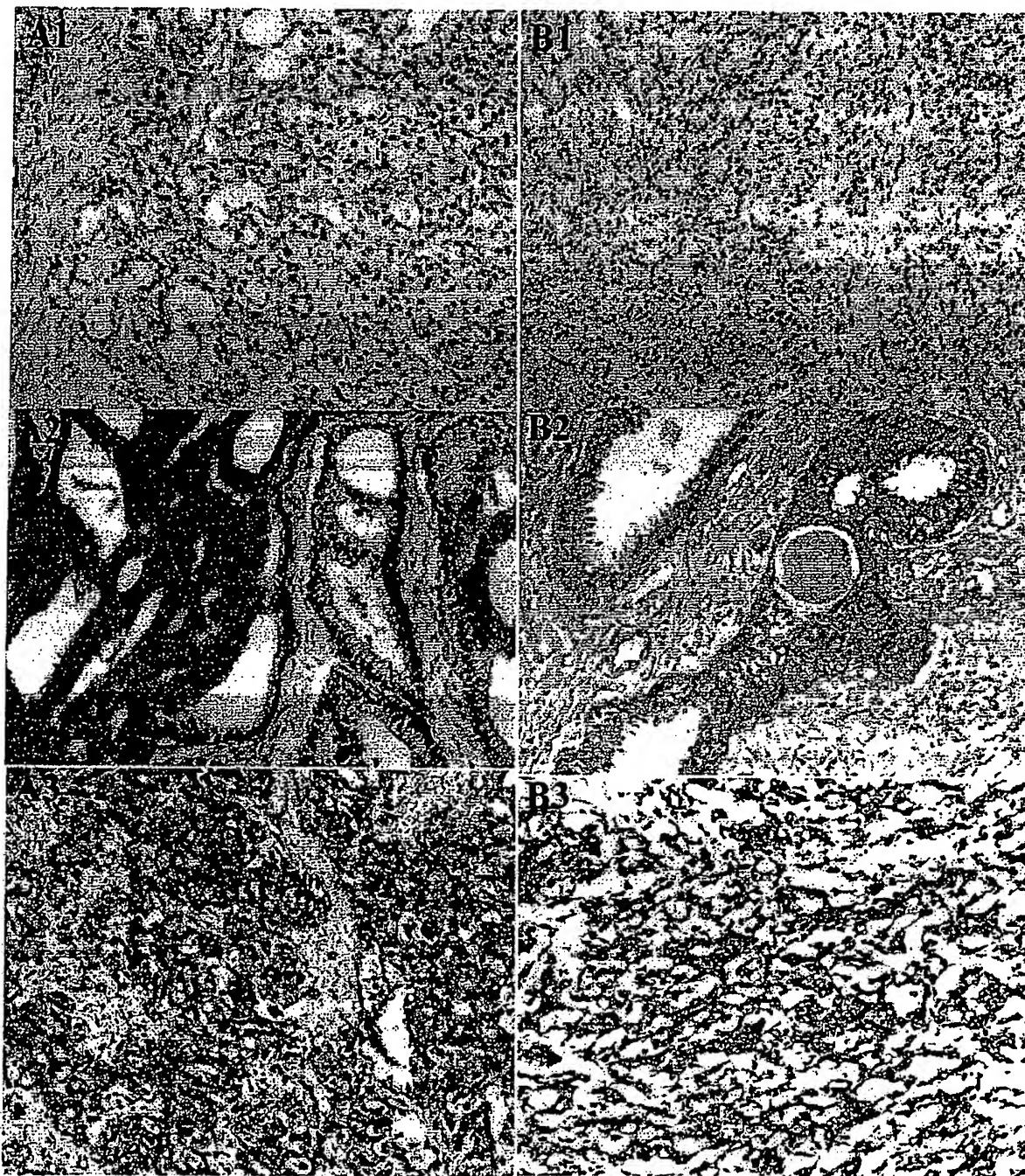


Figure 1

Representatives of PSCA IHC and ISH staining in Pca (A. IHC staining, B. ISH staining, $\times 200$ magnification). A₁, B₁: negative control of IHC and ISH. PBS replacing the primary antibody (A₁) and hybridization with a sense PSCA probe (B₁) showed no background staining. A₂, B₂: a moderately differentiated Pca (Gleason score = $3+3 = 6$) with moderate staining (composite score = 6) in all malignant cells; A₂: IHC shows not only cell surface but also apparent cytoplasmic staining of PSCA protein. A₃, B₃: a poorly differentiated Pca (Gleason score = $4+4 = 8$) with very strong staining (composite score = 9) in all malignant cells.

Discussion

PSCA is homologous to a group of cell surface proteins that mark the earliest phase of hematopoietic development. PSCA mRNA expression is prostate-specific in normal male tissues and is highly up-regulated in both androgen-dependent and-independent Pca xenografts (LAPC-4 tumors). We hypothesize that PSCA may play a role in Pca tumorigenesis and progression, and may serve as a target for Pca diagnosis and treatment. In this study, IHC and ISH showed that in general there were weak or absent PSCA protein and mRNA expression in BPH and low grade PIN tissues. However, PSCA protein and mRNA are widely expressed in HGPIN, the putative precursor of invasive Pca, suggesting that up-regulation of PSCA is an early event in prostate carcinogenesis. Recently, Reiter RE et al [1], using ISH analysis, reported that 97 of 118 (82%) HGPIN specimens stained strongly positive for PSCA mRNA. A very similar finding was seen on mouse PSCA (mPSCA) expression in mouse HGPIN tissues by Tran C. P et al [8]. These data suggest that PSCA may be a new marker associated with transformation of prostate cells and tumorigenesis.

We observed that PSCA protein and mRNA are highly expressed in a large percentage of human prostate cancers, including advanced, poorly differentiated, androgen-independent and metastatic cases. Fluorescence-activated cell sorting and confocal/ immunofluorescent studies demonstrated cell surface expression of PSCA protein in Pca cells [9]. Our IHC expression analysis of PSCA shows not only cell surface but also apparent cytoplasmic staining of PSCA protein in Pca specimens (Figure 1). One possible explanation for this is that anti-PSCA antibody can recognize PSCA peptide precursors that reside in the cytoplasm. Also, it is possible that the positive staining that appears in the cytoplasm is actually from the overlying cell membrane [5]. These data seem to indicate that PSCA is a novel cell surface marker for human Pca.

Our results show that elevated level of PSCA expression correlates with high grade (i.e. poor differentiation), increased tumor stage and progression to androgen-independence of Pca. These findings support the original IHC analyses by Gu Z et al [9], who reported that PSCA protein expressed in 94% of primary Pca and the intensity of PSCA protein expression increased with tumor grade, stage and progression to androgen-independence. Our results also collaborate the recent work of Han KR et al [10], in which the significant association between high PSCA expression and adverse prognostic features such as high Gleason score, seminal vesicle invasion and capsular involvement in Pca was found. It is suggested that PSCA overexpression may be an adverse predictor for recurrence, clinical progression or survival of Pca. Hara H et al [11] used RT-PCR detection of PSA, PSMA and PSCA in 1

ml of peripheral blood to evaluate Pca patients with poor prognosis. The results showed that among 58 Pca patients, each PCR indicated the prognostic value in the hierarchy of PSCA>PSA>PSMA RT-PCR, and extraprostatic cases with positive PSCA PCR indicated lower disease-progression-free survival than those with negative PSCA PCR, demonstrating that PSCA can be used as a prognostic factor. Dubey P et al [12] reported that elevated numbers of PSCA + cells correlate positively with the onset and development of prostate carcinoma over a long time span in the prostates of the TRAMP and PTEN +/- models compared with its normal prostates. Taken together with our present findings, in which PSCA is overexpressed from HGPIN to almost frank carcinoma, it is reasonable and possible to use increased PSCA expression level or increased numbers of PSCA-positive cells in the prostate samples as a prognostic marker to predict the potential onset of this cancer. These data raise the possibility that PSCA may have diagnostic utility or clinical prognostic value in human Pca.

The cause of PSCA overexpression in Pca is not known. One possible mechanism is that it may result from PSCA gene amplification. In humans, PSCA is located on chromosome 8q24.2 [1], which is often amplified in metastatic and recurrent Pca and considered to indicate a poor prognosis [13-15]. Interestingly, PSCA is in close proximity to the c-myc oncogene, which is amplified in >20% of recurrent and metastatic prostate cancers [16,17]. Reiter RE et al [18] reported that PSCA and MYC gene copy numbers were co-amplified in 25% of tumors (five out of twenty), demonstrating that PSCA overexpression is associated with PSCA and MYC coamplification in Pca. Gu Z et al [9] recently reported that in 102 specimens available to compare the results of PSCA immunostaining with their previous mRNA ISH analysis, 92 (90.2%) had identically positive areas of PSCA protein and mRNA expression. Taken together with our findings, in which we detected moderate to strong expression of PSCA protein and mRNA in 34 of 40 (85%) Pca specimens examined simultaneously by IHC and ISH analyses, it is demonstrated that PSCA protein and mRNA overexpressed in human Pca, and that the increased protein level of PSCA was resulted from the upregulated transcription of its mRNA.

At present, the regulation mechanisms of human PSCA expression and its biological function are yet to be elucidated. PSCA expression may be regulated by multiple factors [18]. Watabe T et al [19] reported that transcriptional control is a major component regulating PSCA expression levels. In addition, induction of PSCA expression may be regulated or mediated through cell-cell contact and protein kinase C (PKC) [20]. Homologues of PSCA have diverse activities, and have themselves been involved in

carcinogenesis. Signalling through SCA-2 has been demonstrated to prevent apoptosis in immature thymocytes [21]. Thy-1 is involved in T cell activation and transduces signals through src-like tyrosine kinases [22]. Ly-6 genes have been implicated both in tumorigenesis and in cell-cell adhesion [23-25]. Cell-cell or cell-matrix interaction is critical for local tumor growth and spread to distal sites. From its restricted expression in basal cells of normal prostate and its homology to SCA-2, PSCA may play a role in stem/progenitor cell function, such as self-renewal (i.e. anti-apoptosis) and/or proliferation [1]. Taken together with the results in the present study, we speculate that PSCA may play a role in tumorigenesis and clinical progression of Pca through affecting cell transformation and proliferation. From our results, it is also suggested that PSCA as a new cell surface antigen may have a number of potential uses in the diagnosis, therapy and clinical prognosis of human Pca. PSCA overexpression in prostate biopsies could be used to identify patients at high risk to develop recurrent or metastatic disease, and to discriminate cancers from normal glands in prostatectomy samples. Similarly, the detection of PSCA-overexpressing cells in bone marrow or peripheral blood may identify and predict metastatic progression better than current assays, which identify only PSA-positive or PSMA-positive prostate cells.

In summary, we have shown in this study that PSCA protein and mRNA are maintained in expression from HGPIN through all stages of Pca in a majority of cases, which may be associated with prostate carcinogenesis and correlate positively with high tumor grade (poor cell differentiation), advanced stage and androgen-independent progression. PSCA protein overexpression is due to the upregulation of its mRNA transcription. The results suggest that PSCA may be a promising molecular marker for the clinical prognosis of human Pca and a valuable target for diagnosis and therapy of this tumor.

Competing interests

None declared.

References

- Reiter RE, Gu Z, Watabe T, Thomas G, Szigeti K, David E, Wahl M, Nisitani S, Yamashiro J, Le Beau MM, Loda M, Witte ON: Prostate stem cell antigen: a cell surface marker overexpressed in prostate cancer. *Proc Natl Acad Sci USA* 1998, 95:1735-1740.
- Gleason DF: Histologic grading and clinical staging of prostatic carcinoma. In: *Urologic Pathology: The Prostate* Edited by: Tannebaum M. Philadelphia, Lea & Febiger; 1977:171-197.
- Brawer MK: Prostatic intraepithelial neoplasia: a premalignant lesion. *Hum Pathol* 1992, 23:242-248.
- Amin MB, Ro JY, Ayala AC: Prostatic intraepithelial neoplasia: relationship to adenocarcinoma of prostate. *Pathol Annu* 1994, 29:1-30.
- Amara N, Palapattu GS, Schrage M, Gu Z, Thomas GY, Dorey F, Said J, Reiter RE: Prostate stem cell antigen is overexpressed in human transitional cell carcinoma. *Cancer Res* 2001, 61:4660-4665.
- Hanas JS, Lerner MR, Lightfoot SA, Raczkowski C, Kastens DJ, Brackert DJ, Postier RG: Expression of the cyclin-dependent kinase inhibitor p21 (WAF1/CIP1) and p53 tumor suppressor in dysplastic progression and adenocarcinoma in Barrett esophagus. *Cancer (Phila)* 1999, 86:756-763.
- Egevad L, Gramfors T, Karlberg L: Prognostic value of the Gleason score in prostate cancer. *BJU Int* 2002, 89:538-542.
- Tran CP, Lin C, Yamashiro J, Reiter RE: Prostate stem cell antigen is a marker of late intermediate prostate epithelial cells. *Mol Cancer Res* 2002, 1:113-121.
- Gu Z, Thomas G, Yamashiro J, Shintaku IP, Dorey F, Raitano A, Witte ON, Said JW, Loda M, Reiter RE: Prostate stem cell antigen (PSCA) expression increases with high Gleason score, advanced stage and bone metastasis in prostate cancer. *Oncogene* 2000, 19:1288-1296.
- Han KR, Seligson DB, Liu X, Horvath S, Shintaku PI, Thomas GY, Said JW, Reiter RE: Prostate stem cell antigen expression is associated with gleason score, seminal vesicle invasion and capsular invasion in prostate cancer. *J Urol* 2004, 171:1117-1121.
- Hara H, Kasahara T, Kawasaki T, Bilim V, Obara K, Takahashi K, Tomita Y: Reverse Transcription-Polymerase Chain Reaction Detection of Prostate-specific Antigen, Prostate-specific Membrane Antigen, and Prostate Stem Cell Antigen in One Microliter of Peripheral Blood. *Clin Cancer Res* 2002, 8:1794-1799.
- Dubey P, Wu H, Reiter RE, Witte ON: Alternative pathways to prostate carcinoma activate prostate stem cell antigen expression. *Cancer Res* 2001, 61:3256-3261.
- Visakorpi T, Kallioniemi AH, Syvanen AC, Hyytiäinen ER, Karhu R, Tammela T, Isola JJ, Kallioniemi OP: Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. *Cancer Res* 1995, 55:342-347.
- Sato K, Qian J, Slezak JM, Lieber MM, Bostwick DG, Bergstrahl EJ, Jenkins RB: Clinical significance of alterations of chromosome 8 in high-grade, advanced, nonmetastatic prostate carcinoma. *J Natl Cancer Inst* 1999, 91:1574-1580.
- Van Den Berg C, Guan XY, Von Hoff D, Jenkins R, Bittner J, Griffin C, Kallioniemi O, Visakorpi T, McGill J, Herath J, Epstein J, Sarosdy M, Meltzer P, Trent J: DNA sequence amplification in human prostate cancer identified by chromosome microdissection: potential prognostic implications. *Clin Cancer Res* 1995, 1:11-18.
- Jenkins RB, Qian J, Lieber MM, Bostwick DG: Detection of c-myc oncogene amplification and chromosomal anomalies in metastatic prostatic carcinoma by fluorescence in situ hybridization. *Cancer Res* 1997, 57:524-531.
- Nupponen NN, Kakkola L, Koivisto P, Visakorpi T: Genetic alterations in hormone-refractory recurrent prostate carcinomas. *Am J Pathol* 1998, 153:141-148.
- Reiter RE, Sato I, Thomas G, Qian J, Gu Z, Watabe T, Loda M, Jenkins RB: Coamplification of prostate stem cell antigen (PSCA) and MYC in locally advanced prostate cancer. *Genes Chromosomes Cancer* 2000, 27:95-103.
- Watabe T, Lin M, Donjacour AA, Cunha GR, Witte ON, Reiter RE: Growth, regeneration, and tumorigenesis of the prostate activates the PSCA promoter. *Proc Natl Acad Sci USA* 2002, 99:401-406.
- Bahrenberg G, Brauers A, Joost HG, Jakse G: PSCA expression is regulated by phorbol ester and cell adhesion in the bladder carcinoma cell line RT112. *Cancer Lett* 2001, 168:37-43.
- Noda S, Kosugi A, Saitoh S, Narumiya S, Hamaoka T: Protection from anti-TCR/CD3-induced apoptosis in immature thymocytes by a signal through thymic shared antigen-1/stem cell antigen-2. *J Exp Med* 1996, 183:2355-2360.
- Thomas PM, Samelson LE: The glycosylphosphatidylinositol-anchored Thy-1 molecule interacts with the p60fyn protein tyrosine kinase in T cells. *J Biol Chem* 1992, 267:12317-12322.
- Bamezai A, Rock KL: Overexpressed Ly-6A.2 mediated cell-cell adhesion by binding a ligand expressed on lymphoid cells. *Proc Natl Acad Sci USA* 1995, 92:4294-4298.
- Katz BZ, Eshel R, Sagl-Assif O, Witz IP: An association between high Ly-6A/E expression on tumor cells and a highly malignant phenotype. *Int J Cancer* 1994, 59:684-691.
- Brakenhoff RH, Gerretsen M, Knippels EM, van Dijk M, van Essen H, Weghuis DO, Sinke RJ, Snow GB, van Dongen GA: The human E48 antigen, highly homologous to the murine Ly-6 antigen ThB, is a GPI-anchored molecule apparently involved in keratinocyte cell-cell adhesion. *J Cell Biol* 1995, 129:1677-1689.

Review

Translation Initiation in Cancer: A Novel Target for Therapy¹Funda Meric² and Kelly K. Hunt

Department of Surgical Oncology, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030

Abstract

Translation initiation is regulated in response to nutrient availability and mitogenic stimulation and is coupled with cell cycle progression and cell growth. Several alterations in translational control occur in cancer. Variant mRNA sequences can alter the translational efficiency of individual mRNA molecules, which in turn play a role in cancer biology. Changes in the expression or availability of components of the translational machinery and in the activation of translation through signal transduction pathways can lead to more global changes, such as an increase in the overall rate of protein synthesis and translational activation of the mRNA molecules involved in cell growth and proliferation. We review the basic principles of translational control, the alterations encountered in cancer, and selected therapies targeting translation initiation to help elucidate new therapeutic avenues.

Introduction

The fundamental principle of molecular therapeutics in cancer is to exploit the differences in gene expression between cancer cells and normal cells. With the advent of cDNA array technology, most efforts have concentrated on identifying differences in gene expression at the level of mRNA, which can be attributable either to DNA amplification or to differences in transcription. Gene expression is quite complicated, however, and is also regulated at the level of mRNA stability, mRNA translation, and protein stability.

The power of translational regulation has been best recognized among developmental biologists, because transcription does not occur in early embryogenesis in eukaryotes. For example, in *Xenopus*, the period of transcriptional quiescence continues until the embryo reaches midblastula transition, the 4000-cell stage. Therefore, all necessary mRNA molecules are transcribed during oogenesis and stockpiled in a translationally inactive, masked form. The mRNA are translationally activated at appropriate times during oocyte maturation, fertilization, and

early embryogenesis and thus, are under strict translational control.

Translation has an established role in cell growth. Basically, an increase in protein synthesis occurs as a consequence of mitogenesis. Until recently, however, little was known about the alterations in mRNA translation in cancer, and much is yet to be discovered about their role in the development and progression of cancer. Here we review the basic principles of translational control, the alterations encountered in cancer, and selected therapies targeting translation initiation to elucidate potential new therapeutic avenues.

Basic Principles of Translational Control

Mechanism of Translation Initiation

Translation initiation is the main step in translational regulation. Translation initiation is a complex process in which the initiator tRNA and the 40S and 60S ribosomal subunits are recruited to the 5' end of a mRNA molecule and assembled by eukaryotic translation initiation factors into an 80S ribosome at the start codon of the mRNA (Fig. 1). The 5' end of eukaryotic mRNA is capped, i.e., contains the cap structure m⁷GpppN (7-methyl-guanosine-triphospho-5'-ribonucleoside). Most translation in eukaryotes occurs in a cap-dependent fashion, i.e., the cap is specifically recognized by the eIF4E,³ which binds the 5' cap. The eIF4F translation initiation complex is then formed by the assembly of eIF4E, the RNA helicase eIF4A, and eIF4G, a scaffolding protein that mediates the binding of the 40S ribosomal subunit to the mRNA molecule through interaction with the eIF3 protein present on the 40S ribosome. eIF4A and eIF4B participate in melting the secondary structure of the 5' UTR of the mRNA. The 43S initiation complex (40S/eIF2/Met-tRNA/GTP complex) scans the mRNA in a 5'→3' direction until it encounters an AUG start codon. This start codon is then base-paired to the anticodon of initiator tRNA, forming the 48S initiation complex. The initiation factors are then displaced from the 48S complex, and the 60S ribosome joins to form the 80S ribosome.

Unlike most eukaryotic translation, translation initiation of certain mRNAs, such as the picornavirus RNA, is cap independent and occurs by internal ribosome entry. This mechanism does not require eIF4E. Either the 43S complex can bind the initiation codon directly through interaction with the IRES in the 5' UTR such as in the encephalomyocarditis virus, or it can

Received 5/18/02; revised 7/12/02; accepted 7/22/02.

¹ F. M. is supported by The University of Texas M. D. Anderson Cancer Center Physician-Scientist Program and by NIH Grant 1K08-CA 91895-01. K. K. H. is supported by Department of Defense Award DAMD-17-97-1-7162.

² To whom requests for reprints should be addressed, at Department of Surgical Oncology, Box 444, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030. Phone: (713) 745-4453; Fax: (713) 745-4828; E-mail: fmeric@mdanderson.org.

³ The abbreviations used are: eIF4E, eukaryotic initiation factor 4E; UTR, untranslated region; IRES, internal ribosome entry site; 4E-BP1, eukaryotic initiation factor 4E-binding protein 1; S6K, ribosomal p70 S6 kinase; mTOR, mammalian target of rapamycin; ATM, ataxia telangiectasia mutated; PI3K, phosphatidylinositol 3-kinase; PTEN, phosphatase and tensin homolog deleted from chromosome 10; PP2A, protein phosphatase 2A; TGF- β 3, transforming growth factor- β 3; PAP, poly(A) polymerase; EPA, eicosapentaenoic acid; mda-7, melanoma differentiation-associated gene 7.

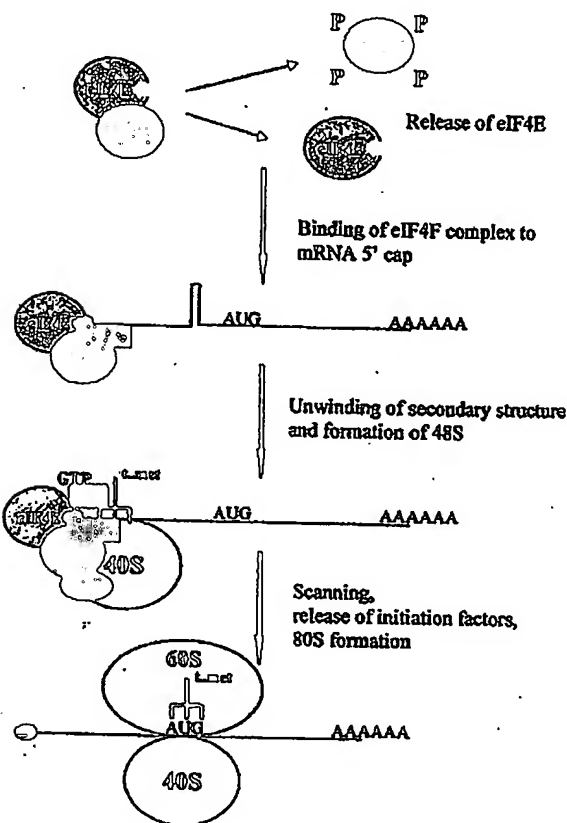


Fig. 1. Translation initiation in eukaryotes. The 4E-BPs are hyperphosphorylated to release eIF4E so that it can interact with the 5' cap, and the eIF4F initiation complex is assembled. The interaction of poly(A) binding protein with the initiation complex and circularization of the mRNA is not depicted in the diagram. The secondary structure of the 5' UTR is melted, the 40S ribosomal subunit is bound to eIF3, and the ternary complex consisting of eIF2, GTP, and the Met-tRNA are recruited to the mRNA. The ribosome scans the mRNA in a 5'→3' direction until an AUG start codon is found in the appropriate sequence context. The initiation factors are released, and the large ribosomal subunit is recruited.

Initially attach to the IRES and then reach the initiation codon by scanning or transfer, as is the case with the poliovirus (1).

Regulation of Translation Initiation

Translation initiation can be regulated by alterations in the expression or phosphorylation status of the various factors involved. Key components in translational regulation that may provide potential therapeutic targets follow.

eIF4E. eIF4E plays a central role in translation regulation. It is the least abundant of the initiation factors and is considered the rate-limiting component for initiation of cap-dependent translation. eIF4E may also be involved in mRNA splicing, mRNA 3' processing, and mRNA nucleocytoplasmic transport (2). eIF4E expression can be increased at the transcriptional level in response to serum or growth factors (3). eIF4E overexpression may cause preferential translation of mRNAs containing excessive secondary structure in their 5' UTR that are normally discriminated against by the trans-

lational machinery and thus are inefficiently translated (4-7). As examples of this, overexpression of eIF4E promotes increased translation of vascular endothelial growth factor, fibroblast growth factor-2, and cyclin D1 (2, 8, 9).

Another mechanism of control is the regulation of eIF4E phosphorylation. eIF4E phosphorylation is mediated by the mitogen-activated protein kinase-interacting kinase 1, which is activated by the mitogen-activated pathway activating extracellular signal-related kinases and the stress-activated pathway acting through p38 mitogen-activated protein kinase (10-13). Several mitogens, such as serum, insulin, angiotensin II, src kinase overexpression, and ras overexpression, lead to eIF4E phosphorylation (14). The phosphorylation status of eIF4E is usually correlated with the translational rate and growth status of the cell; however, eIF4E phosphorylation has also been observed in response to some cellular stresses when translational rates actually decrease (15). Thus, further study is needed to understand the effects of eIF4E phosphorylation on eIF4E activity.

Another mechanism of regulation is the alteration of eIF4E availability by the binding of eIF4E to the eIF4E-binding proteins (4E-BPs, also known as PHAS-I). 4E-BPs compete with eIF4G for a binding site in eIF4E. The binding of eIF4E to the best characterized eIF4E-binding protein, 4E-BP1, is regulated by 4E-BP1 phosphorylation. Hypophosphorylated 4E-BP1 binds to eIF4E, whereas 4E-BP1 hyperphosphorylation decreases this binding. Insulin, angiotensin, epidermal growth factor, platelet-derived growth factor, hepatocyte growth factor, nerve growth factor, insulin-like growth factors I and II, interleukin 3, granulocyte-macrophage colony-stimulating factor + steel factor, gastrin, and the adenovirus have all been reported to induce phosphorylation of 4E-BP1 and to decrease the ability of 4E-BP1 to bind eIF4E (15, 16). Conversely, deprivation of nutrients or growth factors results in 4E-BP1 dephosphorylation, an increase in eIF4E binding, and a decrease in cap-dependent translation.

p70 S6 Kinase. Phosphorylation of ribosomal 40S protein S6 by S6K is thought to play an important role in translational regulation. S6K -/- mouse embryonic cells proliferate more slowly than do parental cells, demonstrating that S6K has a positive influence on cell proliferation (17). S6K regulates the translation of a group of mRNAs possessing a 5' terminal oligopyrimidine tract (5' TOP) found at the 5' UTR of ribosomal protein mRNAs and other mRNAs coding for components of the translational machinery. Phosphorylation of S6K is regulated in part based on the availability of nutrients (18, 19) and is stimulated by several growth factors, such as platelet-derived growth factor and insulin-like growth factor I (20).

eIF2 α Phosphorylation. The binding of the initiator tRNA to the small ribosomal unit is mediated by translation initiation factor eIF2. Phosphorylation of the α -subunit of eIF2 prevents formation of the eIF2/GTP-Met-tRNA complex and inhibits global protein synthesis (21, 22). eIF2 α is phosphorylated under a variety of conditions, such as viral infection, nutrient deprivation, heme deprivation, and apoptosis (22). eIF2 α is phosphorylated by heme-regulated inhibitor, nutrient-regulated protein kinase, and the IFN-induced, double-stranded RNA-activated protein kinase (PKR; Ref. 23).

The mTOR Signaling Pathway. The macrolide antibiotic rapamycin (Sirolimus; Wyeth-Ayeret Research, Collegeville, PA) has been the subject of intensive study because it inhibits signal transduction pathways involved in T-cell activation. The rapamycin-sensitive component of these pathways is mTOR (also called FRAP or RAFT1). mTOR is the mammalian homologue of the yeast TOR proteins that regulate G₁ progression and translation in response to nutrient availability (24). mTOR is a serine-threonine kinase that modulates translation initiation by altering the phosphorylation status of 4E-BP1 and S6K (Fig. 2; Ref. 25).

4E-BP1 is phosphorylated on multiple residues. mTOR phosphorylates the Thr-37 and Thr-46 residues of 4E-BP1 *in vitro* (26); however, phosphorylation at these sites is not associated with a loss of eIF4E binding. Phosphorylation of Thr-37 and Thr-46 is required for subsequent phosphorylation at several COOH-terminal, serum-sensitive sites; a combination of these phosphorylation events appears to be needed to inhibit the binding of 4E-BP1 to eIF4E (25). The product of the ATM gene, p38/MSK1 pathway, and protein kinase C α also play a role in 4E-BP1 phosphorylation (27–29).

S6K and 4E-BP1 are also regulated, in part, by PI3K and its downstream protein kinase Akt. PTEN is a phosphatase that negatively regulates PI3K signaling. PTEN null cells have constitutively active Akt, with increased S6K activity and S6 phosphorylation (30). S6K activity is inhibited both by PI3K inhibitors wortmannin and LY294002 and by mTOR inhibitor rapamycin (24). Akt phosphorylates Ser-2448 in mTOR *in vitro*, and this site is phosphorylated upon Akt activation *in vivo* (31–33). Thus, mTOR is regulated by the PI3K/Akt pathway; however, this does not appear to be the only mode of regulation of mTOR activity. Whether the PI3K pathway also regulates S6K and 4E-BP1 phosphorylation independent of mTOR is controversial.

Interestingly, mTOR autophosphorylation is blocked by wortmannin but not by rapamycin (34). This seeming inconsistency suggests that mTOR-responsive regulation of 4E-BP1 and S6K activity occurs through a mechanism other than intrinsic mTOR kinase activity. An alternate pathway for 4E-BP1 and S6K phosphorylation by mTOR activity is by the inhibition of a phosphatase. Treatment with calyculin A, an inhibitor of phosphatases 1 and 2A, reduces rapamycin-induced dephosphorylation of 4E-BP1 and S6K by rapamycin (35). PP2A interacts with full-length S6K but not with a S6K mutant that is resistant to dephosphorylation resulting from rapamycin. mTOR phosphorylates PP2A *in vitro*; however, how this process alters PP2A activity is not known. These results are consistent with the model that phosphorylation of a phosphatase by mTOR prevents dephosphorylation of 4E-BP1 and S6K, and conversely, that nutrient deprivation and rapamycin block inhibition of the phosphatase by mTOR.

Polyadenylation. The poly(A) tail in eukaryotic mRNA is important in enhancing translation initiation and mRNA stability. Polyadenylation plays a key role in regulating gene expression during oogenesis and early embryogenesis. Some mRNA that are translationally inactive in the oocyte are polyadenylated concomitantly with translational activation in oocyte maturation, whereas other mRNAs that are translationally active during oogenesis are deadenylated and trans-

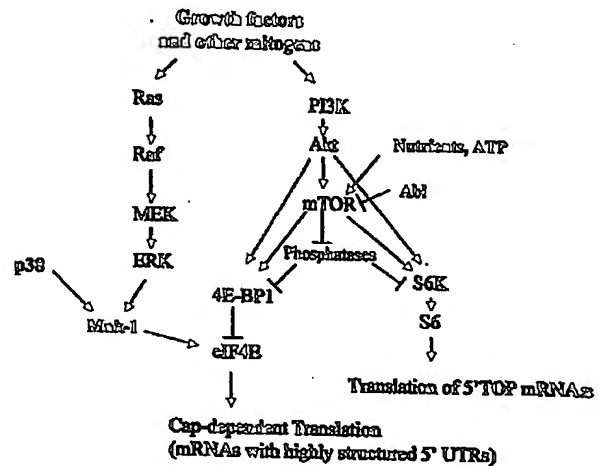


Fig. 2. Regulation of translation initiation by signal transduction pathways. Signaling via p38, extracellular signal-related kinase, PI3K, and mTOR can all activate translation initiation.

lationally silenced (36–38). Thus, control of poly(A) tail synthesis is an important regulatory step in gene expression. The 5' cap and poly(A) tail are thought to function synergistically to regulate mRNA translational efficiency (39, 40).

RNA Packaging. Most RNA-binding proteins are assembled on a transcript at the time of transcription, thus determining the translational fate of the transcript (41). A highly conserved family of Y-box proteins is found in cytoplasmic messenger ribonucleoprotein particles, where the proteins are thought to play a role in restricting the recruitment of mRNA to the translational machinery (41–43). The major mRNA-associated protein, YB-1, destabilizes the interaction of eIF4E and the 5' mRNA cap *in vitro*, and overexpression of YB-1 results in translational repression *in vivo* (44). Thus, alterations in RNA packaging can also play an important role in translational regulation.

Translation Alterations Encountered in Cancer

Three main alterations at the translational level occur in cancer: variations in mRNA sequences that increase or decrease translational efficiency, changes in the expression or availability of components of the translational machinery, and activation of translation through aberrantly activated signal transduction pathways. The first alteration affects the translation of an individual mRNA that may play a role in carcinogenesis. The second and third alterations can lead to more global changes, such as an increase in the overall rate of protein synthesis, and the translational activation of several mRNA species.

Variations in mRNA Sequence

Variations in mRNA sequence affect the translational efficiency of the transcript. A brief description of these variations and examples of each mechanism follow.

Mutations. Mutations in the mRNA sequence, especially in the 5' UTR, can alter its translational efficiency, as seen in the following examples.

c-myc. Saito *et al.* proposed that translation of full-length *c-myc* is repressed, whereas in several Burkitt lymphomas that have deletions of the mRNA 5' UTR, translation of *c-myc* is more efficient (45). More recently, it was reported that the 5' UTR of *c-myc* contains an IRES, and thus *c-myc* translation can be initiated by a cap-independent as well as a cap-dependent mechanism (46, 47). In patients with multiple myeloma, a C→T mutation in the *c-myc* IRES was identified (48) and found to cause an enhanced initiation of translation via internal ribosomal entry (49).

BRCA1. A somatic point mutation (117 G→C) in position -3 with respect to the start codon of the *BRCA1* gene was identified in a highly aggressive sporadic breast cancer (50). Chimeric constructs consisting of the wild-type or mutated *BRCA1* 5' UTR and a downstream luciferase reporter demonstrated a decrease in the translational efficiency with the 5' UTR mutation.

Cyclin-dependent Kinase Inhibitor 2A. Some inherited melanoma kindreds have a G→T transversion at base -34 of cyclin-dependent kinase inhibitor-2A, which encodes a cyclin-dependent kinase 4/cyclin-dependent kinase 6 kinase inhibitor important in G₁ checkpoint regulation (51). This mutation gives rise to a novel AUG translation initiation codon, creating an upstream open reading frame that competes for scanning ribosomes and decreases translation from the wild-type AUG.

Alternate Splicing and Alternate Transcription Start Sites. Alterations in splicing and alternate transcription sites can lead to variations in 5' UTR sequence, length, and secondary structure, ultimately impacting translational efficiency.

ATM. The *ATM* gene has four noncoding exons in its 5' UTR that undergo extensive alternative splicing (52). The contents of 12 different 5' UTRs that show considerable diversity in length and sequence have been identified. These divergent 5' leader sequences play an important role in the translational regulation of the *ATM* gene.

mdm. In a subset of tumors, overexpression of the oncoprotein *mdm2* results in enhanced translation of the *mdm2* mRNA. Use of different promoters leads to two *mdm2* transcripts that differ only in their 5' leaders (53). The longer 5' UTR contains two upstream open reading frames, and this mRNA is loaded with ribosomes inefficiently compared with the short 5' UTR.

BRCA1. In a normal mammary gland, *BRCA1* mRNA is expressed with a shorter leader sequence (5' UTRa), whereas in sporadic breast cancer tissue, *BRCA1* mRNA is expressed with a longer leader sequence (5' UTRb); the translational efficiency of transcripts containing 5' UTRb is 10 times lower than that of transcripts containing 5' UTRa (54).

TGF-β3. *TGF-β3* mRNA includes a 1.1-kb 5' UTR, which exerts an inhibitory effect on translation. Many human breast cancer cell lines contain a novel *TGF-β3* transcript with a 5' UTR that is 870 nucleotides shorter and has a 7-fold greater translational efficiency than the normal *TGF-β3* mRNA (55).

Alternate Polyadenylation Sites. Multiple polyadenylation signals leading to the generation of several transcripts with differing 3' UTR have been described for several mRNA species, such as the *RET* proto-oncogene (56), *ATM* gene (52), tissue inhibitor of metalloproteinases-3 (57), *RHOA*

proto-oncogene (58), and calmodulin-I (59). Although the effect of these alternate 3' UTRs on translation is not yet known, they may be important in RNA-protein interactions that affect translational recruitment. The role of these alterations in cancer development and progression is unknown.

Alterations in the Components of the Translation Machinery

Alterations in the components of translation machinery can take many forms.

Overexpression of eIF4E. Overexpression of eIF4E causes malignant transformation in rodent cells (60) and the deregulation of HeLa cell growth (61). Polunovsky *et al.* (62) found that eIF4E overexpression substitutes for serum and individual growth factors in preserving viability of fibroblasts, which suggests that eIF4E can mediate both proliferative and survival signaling.

Elevated levels of eIF4E mRNA have been found in a broad spectrum of transformed cell lines (63). eIF4E levels are elevated in all ductal carcinoma *in situ* specimens and invasive ductal carcinomas, compared with benign breast specimens evaluated with Western blot analysis (64, 65). Preliminary studies suggest that this overexpression is attributable to gene amplification (66).

There are accumulating data suggesting that eIF4E overexpression can be valuable as a prognostic marker. eIF4E overexpression was found in a retrospective study to be a marker of poor prognosis in stages I to III breast carcinoma (67). Verification of the prognostic value of eIF4E in breast cancer is now under way in a prospective trial (67). However, in a different study, eIF4E expression was correlated with the aggressive behavior of non-Hodgkin's lymphomas (68). In a prospective analysis of patients with head and neck cancer, elevated levels of eIF4E in histologically tumor-free surgical margins predicted a significantly increased risk of local-regional recurrence (9). These results all suggest that eIF4E overexpression can be used to select patients who might benefit from more aggressive systemic therapy. Furthermore, the head and neck cancer data suggest that eIF4E overexpression is a field defect and can be used to guide local therapy.

Alterations in Other Initiation Factors. Alterations in a number of other initiation factors have been associated with cancer. Overproduction of eIF4G, similar to eIF4E, leads to malignant transformation *in vitro* (69). eIF-2α is found in increased levels in bronchioloalveolar carcinomas of the lung (3). Initiation factor eIF-4A1 is overexpressed in melanoma (70) and hepatocellular carcinoma (71). The p40 subunit of translation initiation factor 3 is amplified and overexpressed in breast and prostate cancer (72), and the eIF3-p110 subunit is overexpressed in testicular seminoma (73). The role that overexpression of these initiation factors plays on the development and progression of cancer, if any, is not known.

Overexpression of S6K. S6K is amplified and highly overexpressed in the MCF7 breast cancer cell line, compared with normal mammary epithelium (74). In a study by Barlund *et al.* (74), S6K was amplified in 59 of 668 primary breast tumors, and a statistically significant association was observed between amplification and poor prognosis.

Overexpression of PAP. PAP catalyzes 3' poly(A) synthesis. PAP is overexpressed in human cancer cells compared with normal and virally transformed cells (75). PAP enzymatic activity in breast tumors has been correlated with PAP protein levels (76) and, in mammary tumor cytosols, was found to be an independent factor for predicting survival (76). Little is known, however, about how PAP expression or activity affects the translational profile.

Alterations in RNA-binding Proteins. Even less is known about alterations in RNA packaging in cancer. Increased expression and nuclear localization of the RNA-binding protein YB-1 are indicators of a poor prognosis for breast cancer (77), non-small cell lung cancer (78), and ovarian cancer (79). However, this effect may be mediated at least in part at the level of transcription, because YB-1 increases chemoresistance by enhancing the transcription of a multidrug resistance gene (80).

Activation of Signal Transduction Pathways

Activation of signal transduction pathways by loss of tumor suppressor genes or overexpression of certain tyrosine kinases can contribute to the growth and aggressiveness of tumors. An important mutant in human cancers is the tumor suppressor gene *PTEN*, which leads to the activation of the PI3K/Akt pathway. Activation of PI3K and Akt induces the oncogenic transformation of chicken embryo fibroblasts. The transformed cells show constitutive phosphorylation of S6K and of 4E-BP1 (81). A mutant Akt that retains kinase activity but does not phosphorylate S6K or 4E-BP1 does not transform fibroblasts, which suggests a correlation between the oncogenicity of PI3K and Akt and the phosphorylation of S6K and 4E-BP1 (81).

Several tyrosine kinases such as platelet-derived growth factor, insulin-like growth factor, HER2/neu, and epidermal growth factor receptor are overexpressed in cancer. Because these kinases activate downstream signal transduction pathways known to alter translation initiation, activation of translation is likely to contribute to the growth and aggressiveness of these tumors. Furthermore, the mRNA for many of these kinases themselves are under translational control. For example, HER2/neu mRNA is translationally controlled both by a short upstream open reading frame that represses HER2/neu translation in a cell type-independent manner and by a distinct cell type-dependent mechanism that increases translational efficiency (82). HER2/neu translation is different in transformed and normal cells. Thus, it is possible that alterations at the translational level can in part account for the discrepancy between *HER2/neu* gene amplification detected by fluorescence *in situ* hybridization and protein levels detected by immunohistochemical assays.

Translation Targets of Selected Cancer Therapy

Components of the translation machinery and signal pathways involved in the activation of translation initiation represent good targets for cancer therapy.

Targeting the mTOR Signaling Pathway: Rapamycin and Temsirolumab

Rapamycin inhibits the proliferation of lymphocytes. It was initially developed as an immunosuppressive drug for organ

transplantation. Rapamycin with FKBP 12 (FK506-binding protein, *M_r* 12,000) binds to mTOR to inhibit its function.

Rapamycin causes a small but significant reduction in the initiation rate of protein synthesis (83). It blocks cell growth in part by blocking S6 phosphorylation and selectively suppressing the translation of 5' TOP mRNAs, such as ribosomal proteins, and elongation factors (83–85). Rapamycin also blocks 4E-BP1 phosphorylation and inhibits cap-dependent but not cap-independent translation (17, 86).

The rapamycin-sensitive signal transduction pathway, activated during malignant transformation and cancer progression, is now being studied as a target for cancer therapy (87). Prostate, breast, small cell lung, glioblastoma, melanoma, and T-cell leukemia are among the cancer lines most sensitive to the rapamycin analogue CCI-779 (Wyeth-Ayerst Research; Ref. 87). In rhabdomyosarcoma cell lines, rapamycin is either cytostatic or cytotoxic, depending on the p53 status of the cell; p53 wild-type cells treated with rapamycin arrest in the G₁ phase and maintain their viability, whereas p53 mutant cells accumulate in G₁ and undergo apoptosis (88, 89). In a recently reported study using human primitive neuroectodermal tumor and medulloblastoma models, rapamycin exhibited more cytotoxicity in combination with cisplatin and camptothecin than as a single agent. *In vivo*, CCI-779 delayed growth of xenografts by 160% after 1 week of therapy and 240% after 2 weeks. A single high-dose administration caused a 37% decrease in tumor volume. Growth inhibition *in vivo* was 1.3 times greater, with cisplatin in combination with CCI-779 than with cisplatin alone (90). Thus, preclinical studies suggest that rapamycin analogues are useful as single agents and in combination with chemotherapy.

Rapamycin analogues CCI-779 and RAD001 (Novartis, Basel, Switzerland) are now in clinical trials. Because of the known effect of rapamycin on lymphocyte proliferation, a potential problem with rapamycin analogues is immunosuppression. However, although prolonged immunosuppression can result from rapamycin and CCI-779 administered on continuous-dose schedules, the immunosuppressive effects of rapamycin analogues resolve in ~24 h after therapy (91). The principal toxicities of CCI-779 have included dermatological toxicity, myelosuppression, infection, mucositis, diarrhea, reversible elevations in liver function tests, hyperglycemia, hypokalemia, hypocalcemia, and depression (87, 92–94). Phase II trials of CCI-779 have been conducted in advanced renal cell carcinoma and in stage III/IV breast carcinoma patients who failed with prior chemotherapy. In the results reported in abstract form, although there were no complete responses, partial responses were documented in both renal cell carcinoma and in breast carcinoma (94, 95). Thus, CCI-779 has documented preliminary clinical activity in a previously treated, unselected patient population.

Active investigation is under way into patient selection for mTOR inhibitors. Several studies have found an enhanced efficacy of CCI-779 in PTEN-null tumors (30, 96). Another study found that six of eight breast cancer cell lines were responsive to CCI-779, although only two of these lines lacked PTEN (97). There was, however, a positive correlation between Akt activation and CCI-779 sensitivity (97). This correlation suggests that activation of the PI3K-Akt pathway,

regardless of whether it is attributable to a PTEN mutation or to overexpression of receptor tyrosine kinases, makes cancer cell amenable to mTOR-directed therapy. In contrast, lower levels of the target of mTOR, 4E-BP1, are associated with rapamycin resistance; thus, a lower 4E-BP1/eIF4E ratio may predict rapamycin resistance (98).

Another mode of activity for rapamycin and its analogues appears to be through inhibition of angiogenesis. This activity may be both through direct inhibition of endothelial cell proliferation as a result of mTOR inhibition in these cells or by inhibition of translation of such proangiogenic factors as vascular endothelial growth factor in tumor cells (99, 100).

The angiogenesis inhibitor turostatin, another anticancer drug currently under study, was also found recently to inhibit translation in endothelial cells (101). Through a requisite interaction with Integrin, turostatin inhibits activation of the PI3K/Akt pathway and mTOR in endothelial cells and prevents dissociation of eIF4E from 4E-BP1, thereby inhibiting cap-dependent translation. These findings suggest that endothelial cells are especially sensitive to therapies targeting the mTOR-signaling pathway.

Targeting eIF2 α : EPA, Clotrimazole, mda-7, and Flavonoids

EPA is an n-3 polyunsaturated fatty acid found in the fish-based diets of populations having a low incidence of cancer (102). EPA inhibits the proliferation of cancer cells (103), as well as in animal models (104, 105). It blocks cell division by inhibiting translation initiation (105). EPA releases Ca²⁺ from intracellular stores while inhibiting their refilling, thereby activating PKR. PKR, in turn phosphorylates and inhibits eIF2 α , resulting in the inhibition of protein synthesis at the level of translation initiation. Similarly, clotrimazole, a potent antiproliferative agent *in vitro* and *in vivo*, inhibits cell growth through depletion of Ca²⁺ stores, activation of PKR, and phosphorylation of eIF2 α (106). Consequently, clotrimazole preferentially decreases the expression of cyclins A, E, and D1, resulting in blockage of the cell cycle in G₁.

mda-7 is a novel tumor suppressor gene being developed as a gene therapy agent. Adenoviral transfer of mda-7 (Ad-mda7) induces apoptosis in many cancer cells including breast, colorectal, and lung cancer (107–109). Ad-mda7 also induces and activates PKR, which leads to phosphorylation of eIF2 α and induction of apoptosis (110).

Flavonoids such as genistein and quercetin suppress tumor cell growth. All three mammalian eIF2 α kinases, PKR, heme-regulated inhibitor, and PERK/PEK, are activated by flavonoids, with phosphorylation of eIF2 α and inhibition of protein synthesis (111).

Targeting eIF4A and eIF4E: Antisense RNA and Peptides

Antisense expression of eIF4A decreases the proliferation rate of melanoma cells (112). Sequestration of eIF4E by overexpression of 4E-BP1 is proapoptotic and decreases tumorigenicity (113, 114). Reduction of eIF4E with antisense RNA decreases soft agar growth, increases tumor latency, and increases the rates of tumor doubling times (7). Antisense eIF4E RNA treat-

ment also reduces the expression of angiogenic factors (115) and has been proposed as a potential adjuvant therapy for head and neck cancers, particularly when elevated eIF4E is found in surgical margins. Small molecule inhibitors that bind the eIF4G/eIF4E-binding domain of eIF4E are proapoptotic (116) and are also being actively pursued.

Exploiting Selective Translation for Gene Therapy

A different therapeutic approach that takes advantage of the enhanced cap-dependent translation in cancer cells is the use of gene therapy vectors encoding suicide genes with highly structured 5' UTR. These mRNA would thus be at a competitive disadvantage in normal cells and not translate well, whereas in cancer cells, they would translate more efficiently. For example, the introduction of the 5' UTR of fibroblast growth factor-2 5' to the coding sequence of herpes simplex virus type-1 thymidine kinase gene, allows for selective translation of herpes simplex virus type-1 thymidine kinase gene in breast cancer cell lines compared with normal mammary cell lines and results in selective sensitivity to ganciclovir (117).

Toward the Future

Translation is a crucial process in every cell. However, several alterations in translational control occur in cancer. Cancer cells appear to need an aberrantly activated translational state for survival, thus allowing the targeting of translation initiation with surprisingly low toxicity. Components of the translational machinery, such as eIF4E, and signal transduction pathways involved in translation initiation, such as mTOR, represent promising targets for cancer therapy. Inhibitors of the mTOR have already shown some preliminary activity in clinical trials. It is possible that with the development of better predictive markers and better patient selection, response rates to single-agent therapy can be improved. Similar to other cytostatic agents, however, mTOR inhibitors are most likely to achieve clinical utility in combination therapy. In the interim, our increasing understanding of translation initiation and signal transduction pathways promise to lead to the identification of new therapeutic targets in the near future.

Acknowledgments

We thank Gayle Nesom from The University of Texas M. D. Anderson Cancer Center Department of Scientific Publications for editorial assistance and Dr. Elmer Bernstein for assistance with manuscript preparation.

References

1. Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilpanko, E. V., Shatsky, I. N., Agol, V. I., and Heleen, C. U. Molecular mechanisms of translation initiation in eukaryotes. *Proc. Natl. Acad. Sci. USA*, 98: 7029–7033, 2001.
2. Rosenwald, I. B., Kaspar, R., Rousseau, D., Gehrke, L., Lebouch, P., Chen, J. J., Schmidt, E. V., Sonenberg, N., and London, I. M. Eukaryotic translation initiation factor 4E regulates expression of cyclin D1 at transcriptional and post-transcriptional levels. *J. Biol. Chem.*, 270: 21176–21180, 1995.
3. Rosenwald, I. B., Hutzler, M. J., Wang, S., Savas, L., and Fraire, A. E. Expression of eukaryotic translation initiation factors 4E and 2 α is increased frequently in bronchioloalveolar but not in squamous cell carcinomas of the lung. *Cancer (Phila.)*, 92: 2164–2171, 2001.

4. Darveau, A., Pelletier, J., and Sonenberg, N. Differential efficiencies of *in vitro* translation of mouse c-myc transcripts differing in the 5' untranslated region. *Proc. Natl. Acad. Sci. USA*, 82: 2315-2319, 1985.
5. Kozak, M. Influences of mRNA secondary structure on initiation by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA*, 83: 2850-2854, 1986.
6. Koromilas, A. E., Lazaris-Karatzas, A., and Sonenberg, N. mRNAs containing extensive secondary structure in their 5' non-coding region translate efficiently in cells overexpressing initiation factor eIF-4E. *EMBO J.*, 11: 4153-4158, 1992.
7. Rinker-Schaeffer, C. W., Graff, J. R., De Benedetti, A., Zimmer, S. G., and Rhoads, R. E. Decreasing the level of translation initiation factor 4E with antisense RNA causes reversal of ras-mediated transformation and tumorigenesis of cloned rat embryo fibroblasts. *Int. J. Cancer*, 55: 841-847, 1993.
8. Kevl, C. G., De Benedetti, A., Payne, D. K., Coe, L. L., Laroux, F. S., and Alexander, J. S. Translational regulation of vascular permeability factor by eukaryotic initiation factor 4E: implications for tumor angiogenesis. *Int. J. Cancer*, 65: 785-790, 1998.
9. Nathan, C. A., Franklin, S., Abreo, F. W., Nassar, R., De Benedetti, A., and Glass, J. Analysis of surgical margins with the molecular marker eIF4E: a prognostic factor in patients with head and neck cancer. *J. Clin. Oncol.*, 17: 2809-2814, 1999.
10. Fukunaga, R., and Hunter, T. MINK1, a new MAP kinase-activated protein kinase, isolated by a novel expression screening method for identifying protein kinase substrates. *EMBO J.*, 16: 1921-1933, 1997.
11. Wasikiewicz, A. J., Flynn, A., Proud, C. G., and Cooper, J. A. Mitogen-activated protein kinases activate the serine/threonine kinases Mink1 and Mink2. *EMBO J.*, 16: 1909-1920, 1997.
12. Weng, X., Flynn, A., Wasikiewicz, A. J., Webb, B. L., Vries, R. G., Belnes, I. A., Cooper, J. A., and Proud, C. G. The phosphorylation of eukaryotic initiation factor eIF4E in response to phorbol esters, cell stresses, and cytokines is mediated by distinct MAP kinase pathways. *J. Biol. Chem.*, 273: 9373-9377, 1998.
13. Pyronnet, S., Imataka, H., Gingras, A. C., Fukunaga, R., Hunter, T., and Sonenberg, N. Human eukaryotic translation initiation factor 4G (eIF4G) recruits Mink1 to phosphorylate eIF4E. *EMBO J.*, 18: 270-278, 1999.
14. Klein, M., Schaper, G. C., Voorma, H. O., and Thomas, A. A. Regulation of translation initiation factors by signal transduction. *Eur. J. Biochem.*, 253: 531-544, 1998.
15. Raught, B., and Gingras, A. C. eIF4E activity is regulated at multiple levels. *Int. J. Biochem. Cell Biol.*, 31: 43-57, 1999.
16. Takeuchi, K., Shibamoto, S., Nagamine, K., Shigemori, I., Omura, S., Kitamura, N., and Ito, F. Signaling pathways leading to transcription and translation cooperatively regulate the transient increase in expression of c-Fos protein. *J. Biol. Chem.*, 276: 26077-26083, 2001.
17. Kawasome, H., Papst, P., Webb, S., Keller, G. M., Johnson, G. L., Gelfand, E. W., and Terada, N. Targeted disruption of p70(s6k) defines its role in protein synthesis and rapamycin sensitivity. *Proc. Natl. Acad. Sci. USA*, 95: 6033-6038, 1998.
18. Christie, G. R., Hajdich, E., Hundal, H. S., Proud, C. G., and Taylor, P. M. Intracellular sensing of amino acids in *Xenopus laevis* oocytes stimulates p70 S6 kinase in a target of rapamycin-dependent manner. *J. Biol. Chem.*, 277: 9952-9957, 2002.
19. Hara, K., Yonezawa, K., Weng, Q. P., Kozlowski, M. T., Belham, C., and Avruch, J. Amino acid sufficiency and mTOR regulate p70 S6 kinase and eIF-4E BP1 through a common effector mechanism. *J. Biol. Chem.*, 273: 14484-14494, 1998.
20. Graves, L. M., Bornfeldt, K. E., Argast, G. M., Krebs, E. G., Kong, X., Lin, T. A., and Lawrence, J. C., Jr. cAMP- and rapamycin-sensitive regulation of the association of eukaryotic initiation factor 4E and the translational regulator PHAS-1 in actin smooth muscle cells. *Proc. Natl. Acad. Sci. USA*, 92: 7222-7228, 1995.
21. Merrick, W. C., and Hershey, J. W. B. The pathway and mechanism of eukaryotic protein synthesis. In: J. W. B. Hershey and M. B. Mathews (eds.), *Translational Control*, pp. 31-69. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1998.
22. Kimball, S. R. Eukaryotic initiation factor eIF2. *Int. J. Biochem. Cell Biol.*, 31: 25-29, 1999.
23. Jagus, R., Joshi, B., and Barber, G. N. PKR, apoptosis and cancer. *Int. J. Biochem. Cell Biol.*, 31: 123-138, 1999.
24. Thomas, G., and Hall, M. N. TOR signaling and control of cell growth. *Curr. Opin. Cell Biol.*, 9: 782-787, 1997.
25. Gingras, A. C., Raught, B., and Sonenberg, N. Regulation of translation initiation by FRAP/mTOR. *Genes Dev.*, 16: 807-828, 2001.
26. Gingras, A. C., Gygi, S. P., Raught, B., Polakiewicz, R. D., Abraham, R. T., Hockstra, M. F., Aebersold, R., and Sonenberg, N. Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes Dev.*, 13: 1422-1437, 1999.
27. Kumar, V., Pandey, P., Sabatini, D., Kumar, M., Majumder, P., Bhat, A., Carmichael, G., Kufe, D., and Kharbanda, S. Functional interaction between RAFT1/FRAP/mTOR and protein kinase C δ in the regulation of cap-dependent initiation of translation. *EMBO J.*, 19: 1087-1097, 2000.
28. Yang, D. Q., and Kastan, M. B. Participation of ATM in insulin signaling through phosphorylation of eIF-4E-binding protein 1. *Nat. Cell Biol.*, 2: 893-898, 2000.
29. Liu, G., Zhang, Y., Bode, A. M., Ma, W. Y., and Dong, Z. Phosphorylation of 4E-BP1 is mediated by the p38/MKK1 pathway in response to UVB irradiation. *J. Biol. Chem.*, 277: 8810-8816, 2002.
30. Neshat, M. S., Mellingerhoff, I. K., Tran, C., Stiles, B., Thomas, G., Petersen, R., Frost, P., Gibbons, J. J., Wu, H., and Sawyers, C. L. Enhanced sensitivity of PTEN-deficient tumors to inhibition of FRAP/mTOR. *Proc. Natl. Acad. Sci. USA*, 98: 10314-10319, 2001.
31. Schaeuble, A., Hudson, G. C., Homma, J. L., Yin, P., O'Connell, D. W., Karnitz, L. M., and Abraham, R. T. A direct linkage between the phosphatidylinositol 3-kinase-AKT signaling pathway and the mammalian target of rapamycin in mitogen-stimulated and transformed cells. *Cancer Res.*, 60: 3504-3513, 2000.
32. Scott, P. H., and Lawrence, J. C., Jr. Attenuation of mammalian target of rapamycin activity by increased cAMP in 3T3-L1 adipocytes. *J. Biol. Chem.*, 273: 34498-34501, 1998.
33. Reynolds, I. T., Bodine, S. C., and Lawrence, J. C., Jr. Control of Ser2448 phosphorylation in the mammalian target of rapamycin by insulin and skeletal muscle load. *J. Biol. Chem.*, 277: 17657-17662, 2002.
34. Peterson, R. T., Beal, P. A., Combs, M. J., and Schreiber, S. L. FKBP12-rapamycin-associated protein (FRAP) autophosphorylates at serine 2481 under translationally repressive conditions. *J. Biol. Chem.*, 275: 7416-7423, 2000.
35. Peterson, R. T., Desai, B. N., Hardwick, J. S., and Schreiber, S. L. Protein phosphatase 2A interacts with the 70-kDa S6 kinase and is activated by inhibition of FKBP12-rapamycin-associated protein. *Proc. Natl. Acad. Sci. USA*, 96: 4438-4442, 1999.
36. McGrew, L. L., Dworkin-Rastl, E., Dworkin, M. B., and Richter, J. D. Poly(A) elongation during *Xenopus* oocyte maturation is required for translational recruitment and is mediated by a short sequence element. *Genes Dev.*, 3: 803-815, 1989.
37. Sheets, M. D., Wu, M., and Wickens, M. Polyadenylation of c-mos mRNA as a control point in *Xenopus* meiotic maturation. *Nature (Lond.)*, 374: 511-516, 1995.
38. Varnum, S. M., and Warrington, W. M. Deadenylation of maternal mRNAs during *Xenopus* oocyte maturation does not require specific cis-sequences: a default mechanism for translational control. *Genes Dev.*, 4: 2278-2286, 1990.
39. Gallie, D. R. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev.*, 5: 2108-2116, 1991.
40. Sachs, A. B., and Varani, G. Eukaryotic translation initiation: there are (at least) two sides to every story. *Nat. Struct. Biol.*, 7: 358-361, 2000.
41. Wolffe, A. P., and Merlo, F. Coupling transcription to translation: a novel site for the regulation of eukaryotic gene expression. *Int. J. Biochem. Cell Biol.*, 28: 247-257, 1996.
42. Evdokimova, V. M., Wei, C. L., Sitkov, A. S., Simonenko, P. N., Lazarev, O. A., Vasilenko, K. S., Ustinov, V. A., Hershey, J. W., and Ovchinnikov, L. P. The major protein of messenger ribonucleoprotein particles in somatic cells is a member of the Y-box binding transcription factor family. *J. Biol. Chem.*, 270: 3188-3192, 1995.
43. Matsumoto, K., Merlo, F., and Wolffe, A. P. Translational repression dependent on the interaction of the *Xenopus* Y-box protein FRGY2 with mRNA. Role of the cold shock domain, tail domain, and selective RNA sequence recognition. *J. Biol. Chem.*, 271: 22708-22712, 1996.

44. Evdokimova, V., Ruzanov, P., Imataka, H., Raught, B., Svítidn, Y., Ovchinnikov, L. P., and Sonenberg, N. The major mRNA-associated protein YB-1 is a potent 5' cap-dependent mRNA stabilizer. *EMBO J.*, 20: 5491-5502, 2001.
45. Saito, H., Hayday, A. C., Wiman, K., Hayward, W. S., and Toneyawa, S. Activation of the *c-myc* gene by translocation: a model for translational control. *Proc. Natl. Acad. Sci. USA*, 80: 7476-7480, 1983.
46. Nambu, C., Lafon, L., Audigier, S., Gensac, M. C., Vagner, S., Huez, G., and Prats, A. C. Alternative translation of the proto-oncogene *c-myc* by an internal ribosome entry site. *J. Biol. Chem.*, 272: 32081-32085, 1997.
47. Stonest, M., Paulin, F. E., Le Quesne, J. P., Chappell, S. A., and Willis, A. E. *c-Myc* 5' untranslated region contains an internal ribosome entry segment. *Oncogene*, 16: 423-428, 1998.
48. Paulin, F. E., West, M. J., Sullivan, N. F., Whitney, R. L., Lyne, L., and Willis, A. E. Aberrant translational control of the *c-myc* gene in multiple myeloma. *Oncogene*, 13: 505-513, 1996.
49. Chappell, S. A., LeQuesne, J. P., Paulin, F. E., de Schootmaester, M. L., Stonest, M., Soutar, R. L., Reiston, S. H., Helfrich, M. H., and Willis, A. E. A mutation in the *c-myc*-IRES leads to enhanced internal ribosome entry in multiple myeloma: a novel mechanism of oncogene de-regulation. *Oncogene*, 18: 4437-4440, 2000.
50. Signori, E., Bagni, C., Papa, S., Primerano, B., Rinaldi, M., Amaldi, F., and Fazio, V. M. A somatic mutation in the 5'UTR of *BRCA1* gene in sporadic breast cancer causes down-modulation of translation efficiency. *Oncogene*, 20: 4588-4600, 2001.
51. Liu, L., Diworth, D., Gao, L., Morzon, J., Summers, A., Lessam, N., and Hagg, D. Mutation of the *CDKN2A* 5' UTR creates an aberrant initiation codon and predisposes to melanoma. *Nat. Genet.*, 21: 128-132, 1998.
52. Savitsky, K., Platzer, M., Uziel, T., Glid, S., Sarti, A., Rosenthal, A., Eroy-Stein, O., Shilo, Y., and Rotman, G. Ataxia-telangiectasia: structural diversity of untranslated sequences suggests complex post-transcriptional regulation of *ATM* gene expression. *Nucleic Acids Res.*, 25: 1878-1884, 1997.
53. Brown, C. Y., Mila, G. J., Pineda, M., George, D. L., and Morris, D. R. Role of two upstream open reading frames in the translational control of oncogene *mdm2*. *Oncogene*, 18: 5631-5637, 1999.
54. Sobczak, K., and Krzyzosiak, W. J. Structural determinants of *BRCA1* translational regulation. *J. Biol. Chem.*, 277: 17349-17358, 2002.
55. Artick, B. A., Grendell, R. L., and Griffin, L. A. Enhanced translational efficiency of a novel transforming growth factor β mRNA in human breast cancer cells. *Mol. Cell. Biol.*, 14: 618-628, 1994.
56. Myers, S. M., Eng, C., Ponder, B. A., and Mulligan, L. M. Characterization of *RET* proto-oncogene 3' splicing variants and polyadenylation sites: a novel C-terminus for *RET*. *Oncogene*, 11: 2039-2045, 1995.
57. Byrne, J. A., Tomasello, C., Rouyer, N., Bellocq, J. P., Rio, M. C., and Basset, P. The tissue inhibitor of metalloproteinases-3 gene in breast carcinoma: Identification of multiple polyadenylation sites and a stromal pattern of expression. *Mol. Med.*, 1: 418-427, 1995.
58. Moscow, J. A., He, R., Gudas, J. M., and Cowan, K. H. Utilization of multiple polyadenylation signals in the human *RHOA* protooncogene. *Gene (Amst.)*, 144: 229-236, 1994.
59. Sentenre-Lesentants, S., Alag, A. S., and Sobel, M. E. Multiple mRNA species are generated by alternate polyadenylation from the human *calmodulin-1* gene. *J. Cell. Biochem.*, 58: 445-454, 1995.
60. Lazaris-Karatzas, A., Montina, K. S., and Sonenberg, N. Malignant transformation by a eukaryotic initiation factor subunit that binds to mRNA 5' cap. *Nature (Lond.)*, 345: 544-547, 1990.
61. De Benedetti, A., and Rhoads, R. E. Overexpression of eukaryotic protein synthesis initiation factor 4E in HeLa cells results in aberrant growth and morphology. *Proc. Natl. Acad. Sci. USA*, 87: 8212-8216, 1990.
62. Polunovsky, V. A., Rosenwald, I. B., Tan, A. T., White, J., Chiang, L., Sonenberg, N., and Bittman, P. B. Translational control of programmed cell death: eukaryotic translation initiation factor 4E blocks apoptosis in growth-factor-restricted fibroblasts with physiologically expressed or de-regulated *Myo*. *Mol. Cell. Biol.*, 16: 6573-6581, 1996.
63. Miyagi, Y., Sugiyama, A., Asai, A., Okazaki, T., Kuchino, Y., and Kerr, S. J. Elevated levels of eukaryotic translation initiation factor eIF-4E mRNA in a broad spectrum of transformed cell lines. *Cancer Lett.*, 91: 247-252, 1995.
64. Kerecizta, V., Smiley, K., Hu, B., Smith, A., Gelder, F., and De Benedetti, A. The proto-oncogene/translation factor eIF4E: a survey of its expression in breast carcinomas. *Int. J. Cancer*, 64: 27-31, 1995.
65. Li, B. D., Liu, L., Dawson, M., and De Benedetti, A. Overexpression of eukaryotic initiation factor 4E (eIF4E) in breast carcinoma. *Cancer (Phila.)*, 79: 2385-2390, 1997.
66. Sorrells, D. L., Black, D. R., Meschonat, C., Rhoads, R., De Benedetti, A., Gao, M., Williams, B. J., and Li, B. D. Detection of eIF4E gene amplification in breast cancer by competitive PCR. *Ann. Surg. Oncol.*, 5: 232-237, 1998.
67. Li, B. D., McDonald, J. C., Nassar, R., and De Benedetti, A. Clinical outcome in stage I to III breast carcinoma and eIF4E overexpression. *Ann. Surg.*, 227: 756-761; discussion, 761-763, 1998.
68. Wang, S., Rosenwald, I. B., Hutzler, M. J., Pihan, G. A., Savas, L., Chen, J. J., and Woda, B. A. Expression of the eukaryotic translation initiation factors 4E and 2a in non-Hodgkin's lymphomas. *Am. J. Pathol.*, 155: 247-255, 1999.
69. Futuchi-Shimogori, T., Ishii, I., Kashiwagi, K., Mashiba, H., Edamoto, H., and Igarashi, K. Malignant transformation by overproduction of translation initiation factor eIF4G. *Cancer Res.*, 57: 5041-5044, 1997.
70. Eberle, J., Krasagakis, K., and Orfanos, C. E. Translation initiation factor eIF-4A1 mRNA is consistently overexpressed in human melanoma cells *in vitro*. *Int. J. Cancer*, 77: 398-401, 1997.
71. Shuda, M., Kondoh, N., Tanaka, K., Ryo, A., Watanabe, Y., Hada, A., Goseki, M., Igarashi, T., Matsuda, K., Aihara, T., Horuchi, S., Saitoh, M., Yamamoto, N., and Yamamoto, M. Enhanced expression of translation factor mRNAs in hepatocellular carcinoma. *Anticancer Res.*, 20: 2489-2494, 2000.
72. Nuppenen, M. M., Portila, K., Kakkola, L., Tanner, M., Persson, K., Borg, A., Isola, J., and Vastorp, T. Amplification and overexpression of p40 subunit of eukaryotic translation initiation factor 3 in breast and prostate cancer. *Am. J. Pathol.*, 154: 1777-1783, 1999.
73. Roth, M., Ko, Y., Albani, P., and Werner, N. Eukaryotic initiation factor 3 p110 mRNA is overexpressed in testicular seminomas. *Am. J. Pathol.*, 157: 1597-1604, 2000.
74. Barlund, M., Faroozan, F., Kononen, J., Bubendorf, L., Chen, Y., Bittner, M. L., Torst, J., Haas, P., Bucher, C., Sauter, G., Kallioniemi, O. P., and Kallioniemi, A. Detecting activation of ribosomal protein S6 kinase by complementary DNA and tissue microarray analysis. *J. Natl. Cancer Inst.* (Bethesda), 92: 1252-1259, 2000.
75. Topalian, S. L., Kaneko, S., Gonzales, M. I., Bond, G. L., Ward, Y., and Manley, J. L. Identification and functional characterization of neo-poly(A) polymerase, an RNA processing enzyme overexpressed in human tumors. *Mol. Cell. Biol.*, 21: 5314-5323, 2001.
76. Scourlas, A., Tafari, M., Ardavanis, A., Courtis, N., Dimitriadis, E., Yotis, J., Tsipalis, C. M., and Tragas, T. Polyadenylate polymerase enzymatic activity in mammary tumor cytosols: a new independent prognostic marker in primary breast cancer. *Cancer Res.*, 60: 5427-5433, 2000.
77. Janz, M., Harbeck, N., Dittmar, P., Berger, U., Schmidt, A., Juchott, K., Schmitt, M., and Royer, H. D. Y-box factor YB-1 predicts drug resistance and patient outcome in breast cancer independent of clinically relevant tumor biologic factors HER2, uPA and PAI-1. *Int. J. Cancer*, 97: 278-282, 2002.
78. Shibahara, K., Sugio, K., Osaki, T., Uchiyama, T., Maehara, Y., Kohno, K., Yasumoto, K., Sugimachi, K., and Kuwano, M. Nuclear expression of the Y-box binding protein, YB-1, as a novel marker of disease progression in non-small cell lung cancer. *Clin. Cancer Res.*, 7: 3151-3155, 2001.
79. Kamura, T., Yahara, H., Amada, S., Ogawa, S., Sonoda, T., Kobayashi, H., Mitsumoto, M., Kohno, K., Kuwano, M., and Nakano, H. Is nuclear expression of Y box-binding protein-1 a new prognostic factor in ovarian serous adenocarcinoma? *Cancer (Phila.)*, 85: 2450-2454, 1999.
80. Bargou, R. C., Juchott, K., Wagener, C., Bergmann, S., Metzner, S., Bonnmert, K., Mapara, M. Y., Winzer, K. J., Dietel, M., Dorken, B., and Royer, H. D. Nuclear localization and increased levels of transcription factor YB-1 in primary human breast cancers are associated with intraluc *MDR1* gene expression. *Nat. Med.*, 3: 447-450, 1997.
81. Aoki, M., Blazek, E., and Vogt, P. K. A role of the kinase mTOR in cellular transformation induced by the oncoproteins P3k and Akt. *Proc. Natl. Acad. Sci. USA*, 98: 136-141, 2001.
82. Child, S. J., Miller, M. K., and Geballe, A. P. Cell type-dependent and -independent control of HER-2/neu translation. *Int. J. Biochem. Cell Biol.*, 37: 201-213, 1999.

83. Jefferies, H. B., Reinhard, C., Kozma, S. C., and Thomas, G. Rapamycin selectively represses translation of the "polypyrimidine tract" mRNA family. *Proc. Natl. Acad. Sci. USA*, 91: 4441-4445, 1994.
84. Terada, N., Patel, H. R., Takase, K., Kohno, K., Naim, A. C., and Gelfand, E. W. Rapamycin selectively inhibits translation of mRNAs encoding elongation factors and ribosomal proteins. *Proc. Natl. Acad. Sci. USA*, 91: 11477-11481, 1994.
85. Jefferies, H. B., Furnagall, S., Dennis, P. B., Reinhard, C., Pearson, R. B., and Thomas, G. Rapamycin suppresses 5'TOP mRNA translation through inhibition of p70S6K. *EMBO J.*, 18: 3693-3704, 1999.
86. Beretta, L., Gingras, A. C., Svitkin, Y. V., Hall, M. N., and Sonenberg, N. Rapamycin blocks the phosphorylation of 4E-BP1 and inhibits cap-dependent initiation of translation. *EMBO J.*, 15: 658-664, 1996.
87. Hidalgo, M., and Rowinsky, E. K. The rapamycin-sensitive signal transduction pathway as a target for cancer therapy. *Oncogene*, 19: 6680-6688, 2000.
88. Hosoi, H., Dilling, M. B., Shikata, T., Liu, L. N., Shi, L., Ashmun, R. A., Germann, G. S., Abraham, R. T., and Houghton, P. J. Rapamycin causes poorly reversible inhibition of mTOR and induces p53-independent apoptosis in human rhabdomyosarcoma cells. *Cancer Res.*, 63: 888-894, 1999.
89. Huang, S., and Houghton, P. J. Resistance to rapamycin: a novel anticancer drug. *Cancer Metastasis Rev.*, 20: 69-78, 2001.
90. Georger, B., Kerr, K., Teng, C. B., Fung, K. M., Powell, B., Sutton, L. M., Phillips, P. C., and Jansa, A. J. Antitumor activity of the rapamycin analog CCI-779 in human primitive neuroectodermal tumor/mesodermoblastoma models as single agent and in combination chemotherapy. *Cancer Res.*, 61: 1527-1532, 2001.
91. Gibbons, J. J., Discafani, C., Peterson, R., Hernandez, R., Skotnick, J., and Frost, P. The effect of CCI-779, a novel macrocyclic anti-tumor agent, on the growth of human tumor cells *in vitro* and in nude mouse xenografts *in vivo*. *Proc. Am. Assoc. Cancer Res.*, 40: 301, 1999.
92. Hidalgo, M., Rowinsky, E., Erlichman, C., Marshall, B., Marks, R., Edwards, T., and Buckner, J. J. A Phase I and pharmacological study of CCI-779 cycle inhibitor. *Ann. Oncol.*, 11 (Suppl. 4): 133, 2001.
93. Alexandra, J., Raymond, E., Depenbrock, H., Melchaldi, S., Angevin, E., Pallet, C., Hanauske, A., Fritsch, J., Feussner, A., and Armand, J. P. CCI-779, a new rapamycin analog, has antitumor activity at doses inducing only mild cutaneous effects and mucositis: early results of an ongoing Phase I study. *Proceedings of the 1999 AACR-NCI-EORTC International Conference, Clin. Cancer Res.*, 5 (Suppl.): 3730a, 1999.
94. Chan, S., Johnston, S., Schauten, M. E., Mross, K., Morant, A., Lehr, A., Feussner, A., Berger, M., and Kirsch, T. First report: a Phase 2 study of the safety and activity of CCI-779 for patients with locally advanced or metastatic breast cancer failing prior chemotherapy. *Proc. Am. Soc. Clin. Oncol.*, 21: 44a, 2002.
95. Atkins, M. B., Hidalgo, M., Stadler, W., Logan, T., Dutcher, J. P., Hudes, G., Park, Y., Marshall, B., Bori, J., and Dukart, G. A randomized double-blind Phase 2 study of intravenous CCI-779 administered weekly to patients with advanced renal cell carcinoma. *Proc. Am. Soc. Clin. Oncol.*, 21: 10a, 2002.
96. Smith, S. G., Trinh, C. M., Inge, L. J., Thomas, G., Cloughsey, T. F., Sawyers, C. L., and Mischel, P. S. PTEN expression status predicts glioblastoma cell sensitivity to CCI-779. *Proc. Am. Assoc. Cancer Res.*, 43: 335, 2002.
97. Yu, K., Toni-Berza, L., Discafani, C., Zhang, W. G., Skotnick, J., Frost, P., and Gibbons, J. J. mTOR, a novel target in breast cancer: the effect of CCI-779, an mTOR inhibitor, in preclinical models of breast cancer. *Endocr. Relat. Cancer*, 8: 249-258, 2001.
98. Dilling, M. B., Germann, G. S., Dudkin, L., Jayaraman, A. L., Zhang, X., Harwood, F. C., and Houghton, P. J. 4E-binding proteins, the suppressors of eukaryotic initiation factor 4E, are downregulated in cells with acquired or intrinsic resistance to rapamycin. *J. Biol. Chem.*, 277: 13907-13917, 2002.
99. Guba, M., von Breitenbuch, P., Steinbauer, M., Koehl, G., Fiegel, S., Hornung, M., Bruns, C. J., Zuelke, C., Farkas, S., Anthuber, M., Jauch, K. W., and Geisler, E. K. Rapamycin inhibits primary and metastatic tumor growth by antiangiogenesis: Involvement of vascular endothelial growth factor. *Nat. Med.*, 8: 128-135, 2002.
100. Lane, H. A., Schell, C., Theuer, A., O'Reilly, T., and Wood, J. Anti-angiogenic activity of RAD001, an orally active anticancer agent. *Proc. Am. Assoc. Cancer Res.*, 43: 184, 2002.
101. Maeshima, Y., Sudhaker, A., Lively, J. C., Ueki, K., Kharbanda, S., Kahn, C. R., Sonenberg, N., Hynes, R. O., and Kalluri, R. Tumstatin, an endothelial cell-specific inhibitor of protein synthesis. *Science (Wash. DC)*, 295: 140-143, 2002.
102. Caygill, C. P., Charlett, A., and Hill, M. J. Fat, fish, fish oil and cancer. *Br. J. Cancer*, 74: 159-164, 1996.
103. Falconer, J. S., Ross, J. A., Fearon, K. C., Hawkins, R. A., O'Riordan, M. G., and Carter, D. C. Effect of eicosapentaenoic acid and other fatty acids on the growth *in vitro* of human pancreatic cancer cell lines. *Br. J. Cancer*, 69: 828-832, 1994.
104. Noguchi, M., Minami, M., Yagasaki, R., Kinoshita, K., Earashi, M., Kikawa, H., Taniya, T., and Miyazaki, I. Chemoprevention of DMBA-induced mammary carcinogenesis in rats by low-dose EPA and DHA. *Br. J. Cancer*, 75: 348-353, 1997.
105. Palakurthi, S. S., Fluckiger, R., Aktas, H., Changolkar, A. K., Shahsafaee, A., Harnett, S., Kille, E., and Halperin, J. A. Inhibition of translation initiation mediates the anticancer effect of the n-3 polyunsaturated fatty acid eicosapentaenoic acid. *Cancer Res.*, 60: 2919-2925, 2000.
106. Aktas, H., Fluckiger, R., Acosta, J. A., Savaga, J. M., Palakurthi, S. S., and Halperin, J. A. Depletion of intracellular Ca^{2+} stores, phosphorylation of eIF2 α , and sustained inhibition of translation initiation mediate the anticancer effects of clotrimazole. *Proc. Natl. Acad. Sci. USA*, 95: 8280-8285, 1998.
107. Mhashilkar, A. M., Schrock, R. D., Hindi, M., Liao, J., Siegel, K., Kourouma, F., Zou-Yang, X. H., Onishi, E., Taki, O., Vedvick, T. S., Fanger, G., Stewart, L., Watson, G. J., Smey, D., Fisher, P. B., Seak, T., Roth, J. A., Ramesh, R., and Chada, S. Melanoma differentiation associated gene-7 (*mda-7*): a novel anti-tumor gene for cancer gene therapy. *Mol. Med.*, 7: 271-282, 2001.
108. Su, Z. Z., Madhreddi, M. T., Lin, J. J., Young, C. S., Kitada, S., Reed, J. C., Goldstein, N. I., and Fisher, P. B. The cancer growth suppressor gene *mda-7* selectively induces apoptosis in human breast cancer cells and inhibits tumor growth in nude mice. *Proc. Natl. Acad. Sci. USA*, 95: 14400-14405, 1998.
109. Seak, T., Mhashilkar, A., Chada, S., Branch, C., Roth, J. A., and Ramesh, R. Tumor-suppressive effects by adenovirus-mediated *mda-7* gene transfer in non-small cell lung cancer cell *in vitro*. *Gene Ther.*, 7: 2051-2057, 2000.
110. Pataer, A., Vortuburger, S. A., Barber, G. N., Chada, S., Mhashilkar, A. M., Zou-Yang, H., Stewart, A. L., Belachandran, S., Roth, J. A., Hunt, K. K., and Swisher, S. G. Adenoviral transfer of the melanoma differentiation-associated gene 7 (*mda7*) induces apoptosis of lung cancer cells via up-regulation of the double-stranded RNA-dependent protein kinase (PKR). *Cancer Res.*, 62: 2239-2243, 2002.
111. Ito, T., Wamken, S. P., and May, W. S. Protein synthesis inhibition by flavonoids: roles of eukaryotic initiation factor 2 α kinases. *Biochem. Biophys. Res. Commun.*, 265: 589-594, 1999.
112. Eberle, J., Fecker, L. F., Bittner, J. U., Orfanos, C. E., and Geilen, C. C. Decreased proliferation of human melanoma cell lines caused by antisense RNA against translation factor eIF-4A1. *Br. J. Cancer*, 88: 1957-1962, 2002.
113. Polunovsky, V. A., Gingras, A. C., Sonenberg, N., Peterson, M., Ten, A., Rubins, J. B., Manivel, J. C., and Bitterman, P. B. Translational control of the antiapoptotic function of Ras. *J. Biol. Chem.*, 275: 24776-24780, 2000.
114. D'Cunha, J., Kratzke, M. G., Alter, M. D., Polunovsky, V. A., Bitterman, P. B., and Kratzke, R. A. Over-expression of the translational repressor 4E-BP1 inhibits NSCLC tumorigenicity *in vivo*. *Proc. Am. Assoc. Cancer Res.*, 43: 818-817, 2002.
115. DeFatta, R. J., Nathan, C. A., and De Benedetti, A. Antisense RNA to eIF4E suppresses oncogenic properties of a head and neck squamous cell carcinoma cell line. *Laryngoscope*, 110: 928-933, 2000.
116. Herbert, T. P., Fahrhaus, R., Prescott, A., Lane, D. P., and Proud, C. G. Rapid induction of apoptosis mediated by peptides that bind initiation factor eIF4E. *Curr. Biol.*, 10: 793-798, 2000.
117. DeFatta, R. J., Li, Y., and De Benedetti, A. Selective killing of cancer cells based on translational control of a suicide gene. *Cancer Gene Ther.*, 9: 573-578, 2002.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant : Ashkenazi et al.
App. No. : 09/903,925
Filed : July 11, 2001
For : SECRETED AND
TRANSMEMBRANE
POLYPEPTIDES AND NUCLEIC
ACIDS ENCODING THE SAME

Examiner : Hamud, Fozia M

Group Art Unit 1647

CERTIFICATE OF EXPRESS MAILING

I hereby certify that this correspondence is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to Commissioner of Patents, Washington D.C. 20231 on:

(Date)

Commissioner of Patents
P.O. Box 1450
Alexandria, VA 22313-1450

DECLARATION OF AVI ASHKENAZI, Ph.D UNDER 37 C.F.R. § 1.132

I, Avi Ashkenazi, Ph.D. declare and say as follows: -

1. I am Director and Staff Scientist at the Molecular Oncology Department of Genentech, Inc., South San Francisco, CA 94080.
2. I joined Genentech in 1988 as a postdoctoral fellow. Since then, I have investigated a variety of cellular signal transduction mechanisms, including apoptosis, and have developed technologies to modulate such mechanisms as a means of therapeutic intervention in cancer and autoimmune disease. I am currently involved in the investigation of a series of secreted proteins over-expressed in tumors, with the aim to identify useful targets for the development of therapeutic antibodies for cancer treatment.
3. My scientific Curriculum Vitae, including my list of publications, is attached to and forms part of this Declaration (Exhibit A).
4. Gene amplification is a process in which chromosomes undergo changes to contain multiple copies of certain genes that normally exist as a single copy, and is an important factor in the pathophysiology of cancer. Amplification of certain genes (e.g., Myc or Her2/Neu)

gives cancer cells a growth or survival advantage relative to normal cells, and might also provide a mechanism of tumor cell resistance to chemotherapy or radiotherapy.

5. If gene amplification results in over-expression of the mRNA and the corresponding gene product, then it identifies that gene product as a promising target for cancer therapy, for example by the therapeutic antibody approach. Even in the absence of over-expression of the gene product, amplification of a cancer marker gene - as detected, for example, by the reverse transcriptase TaqMan[®] PCR or the fluorescence *in situ* hybridization (FISH) assays - is useful in the diagnosis or classification of cancer, or in predicting or monitoring the efficacy of cancer therapy. An increase in gene copy number can result not only from intrachromosomal changes but also from chromosomal aneuploidy. It is important to understand that detection of gene amplification can be used for cancer diagnosis even if the determination includes measurement of chromosomal aneuploidy. Indeed, as long as a significant difference relative to normal tissue is detected, it is irrelevant if the signal originates from an increase in the number of gene copies per chromosome and/or an abnormal number of chromosomes.

6. I understand that according to the Patent Office, absent data demonstrating that the increased copy number of a gene in certain types of cancer leads to increased expression of its product, gene amplification data are insufficient to provide substantial utility or well established utility for the gene product (the encoded polypeptide), or an antibody specifically binding the encoded polypeptide. However, even when amplification of a cancer marker gene does not result in significant over-expression of the corresponding gene product, this very absence of gene product over-expression still provides significant information for cancer diagnosis and treatment. Thus, if over-expression of the gene product does not parallel gene amplification in certain tumor types but does so in others, then parallel monitoring of gene amplification and gene product over-expression enables more accurate tumor classification and hence better determination of suitable therapy. In addition, absence of over-expression is crucial information for the practicing clinician. If a gene is amplified but the corresponding gene product is not over-expressed, the clinician accordingly will decide not to treat a patient with agents that target that gene product.

7. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information or belief are believed to be true, and further that these statements were made with the knowledge that willful false statements and the like so

made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful statements may jeopardize the validity of the application or any patent issued thereon.

By: Avi Ashkenazi
Avi Ashkenazi, Ph.D.

Date: 9/15/03

SV 455281 v1
9/12/03 3:06 PM (39780.7000)

EXHIBIT A

CURRICULUM VITAE

Avi Ashkenazi

July 2003

Personal:

Date of birth: 29 November, 1956
Address: 1456 Tarrytown Street, San Mateo, CA 94402
Phone: (650) 578-9199 (home); (650) 225-1853 (office)
Fax: (650) 225-6443 (office)
Email: aa@gene.com

Education:

1983: B.S. in Biochemistry, with honors, Hebrew University, Israel
1986: Ph.D. in Biochemistry, Hebrew University, Israel

Employment:

1983-1986: Teaching assistant, undergraduate level course in Biochemistry
1985-1986: Teaching assistant, graduate level course on Signal Transduction
1986 - 1988: Postdoctoral fellow, Hormone Research Dept., UCSF, and
Developmental Biology Dept., Genentech, Inc., with J. Ramachandran
1988 - 1989: Postdoctoral fellow, Molecular Biology Dept., Genentech, Inc.,
with D. Capon
1989 - 1993: Scientist, Molecular Biology Dept., Genentech, Inc.
1994 -1996: Senior Scientist, Molecular Oncology Dept., Genentech, Inc.
1996-1997: Senior Scientist and Interim director, Molecular Oncology Dept.,
Genentech, Inc.
1997-1990: Senior Scientist and preclinical project team leader, Genentech, Inc.
1999 -2002: Staff Scientist in Molecular Oncology, Genentech, Inc.
2002-present: Staff Scientist and Director in Molecular Oncology, Genentech, Inc.

Awards:

1988: First prize, The Boehringer Ingelheim Award

Editorial:

Editorial Board Member: Current Biology

Associate Editor, Clinical Cancer Research.

Associate Editor, Cancer Biology and Therapy.

Refereed papers:

1. Gertler, A., Ashkenazi, A., and Madar, Z. Binding sites for human growth hormone and ovine and bovine prolactins in the mammary gland and liver of the lactating cow. *Mol. Cell. Endocrinol.* **34**, 51-57 (1984).
2. Gertler, A., Shamay, A., Cohen, N., Ashkenazi, A., Friesen, H., Levanon, A., Gorecki, M., Aviv, H., Hadari, D., and Vogel, T. Inhibition of lactogenic activities of ovine prolactin and human growth hormone (hGH) by a novel form of a modified recombinant hGH. *Endocrinology* **118**, 720-726 (1986).
3. Ashkenazi, A., Madar, Z., and Gertler, A. Partial purification and characterization of bovine mammary gland prolactin receptor. *Mol. Cell. Endocrinol.* **50**, 79-87 (1987).
4. Ashkenazi, A., Pines, M., and Gertler, A. Down-regulation of lactogenic hormone receptors in Nb2 lymphoma cells by cholera toxin. *Biochemistry Internat.* **14**, 1065-1072 (1987).
5. Ashkenazi, A., Cohen, R., and Gertler, A. Characterization of lactogen receptors in lactogenic hormone-dependent and independent Nb2 lymphoma cell lines. *FEBS Lett.* **210**, 51-55 (1987).
6. Ashkenazi, A., Vogel, T., Barash, I., Hadari, D., Levanon, A., Gorecki, M., and Gertler, A. Comparative study on-in-vitro and in-vivo modulation of lactogenic and somatotrophic receptors by native human growth hormone and its modified recombinant analog. *Endocrinology* **121**, 414-419 (1987).
7. Peralta, E., Winslow, J., Peterson, G., Smith, D., Ashkenazi, A., Ramachandran, J., Schimerlik, M., and Capon, D. Primary structure and biochemical properties of an M2 muscarinic receptor. *Science* **236**, 600-605 (1987).
8. Peralta, E., Ashkenazi, A., Winslow, J., Smith, D., Ramachandran, J., and Capon, D. J. Distinct primary structures, ligand-binding properties and tissue-specific expression of four human muscarinic acetylcholine receptors. *EMBO J.* **6**, 3923-3929 (1987).
9. Ashkenazi, A., Winslow, J., Peralta, E., Peterson, G., Schimerlik, M., Capon, D., and Ramachandran, J. An M2 muscarinic receptor subtype coupled to both adenylyl cyclase and phosphoinositide turnover. *Science* **238**, 672-675 (1987).

10. Pines, M., Ashkenazi, A., Cohen-Chapnik, N., Binder, L., and Gertler, A. Inhibition of the proliferation of Nb2 lymphoma cells by femtomolar concentrations of cholera toxin and partial reversal of the effect by 12-o-tetradecanoyl-phorbol-13-acetate. *J. Cell. Biochem.* 37, 119-129 (1988).
11. Peralta, E., Ashkenazi, A., Winslow, J., Ramachandran, J., and Capon, D. Differential regulation of PI hydrolysis and adenylyl cyclase by muscarinic receptor subtypes. *Nature* 334, 434-437 (1988).
12. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functionally distinct G proteins couple different receptors to PI hydrolysis in the same cell. *Cell* 56, 487-493 (1989).
13. Ashkenazi, A., Ramachandran, J., and Capon, D. Acetylcholine analogue stimulates DNA synthesis in brain-derived cells via specific muscarinic acetylcholine receptor subtypes. *Nature* 340, 146-150 (1989).
14. Lammare, D., Ashkenazi, A., Fleury, S., Smith, D., Sekaly, R., and Capon, D. The MHC-binding and gp120-binding domains of CD4 are distinct and separable. *Science* 245, 743-745 (1989).
15. Ashkenazi, A., Presta, L., Marsters, S., Camerato, T., Rosenthal, K., Fendly, B., and Capon, D. Mapping the CD4 binding site for human immunodeficiency virus type 1 by alanine-scanning mutagenesis. *Proc. Natl. Acad. Sci. USA.* 87, 7150-7154 (1990).
16. Chamow, S., Peers, D., Byrn, R., Mulkerrin, M., Harris, R., Wang, W., Bjorkman, P., Capon, D., and Ashkenazi, A. Enzymatic cleavage of a CD4 immunoadhesin generates crystallizable, biologically active Fd-like fragments. *Biochemistry* 29, 9885-9891 (1990).
17. Ashkenazi, A., Smith, D., Marsters, S., Riddle, L., Gregory, T., Ho, D., and Capon, D. Resistance of primary isolates of human immunodeficiency virus type 1 to soluble CD4 is independent of CD4-rgp120 binding affinity. *Proc. Natl. Acad. Sci. USA.* 88, 7056-7060 (1991).
18. Ashkenazi, A., Marsters, S., Capon, D., Chamow, S., Figari, I., Pennica, D., Goeddel, D., Palladino, M., and Smith, D. Protection against endotoxic shock by a tumor necrosis factor receptor immunoadhesin. *Proc. Natl. Acad. Sci. USA.* 88, 10535-10539 (1991).
19. Moore, J., McKeating, J., Huang, Y., Ashkenazi, A., and Ho, D. Virions of primary HIV-1 isolates resistant to sCD4 neutralization differ in sCD4 affinity and glycoprotein gp120 retention from sCD4-sensitive isolates. *J. Virol.* 66, 235-243 (1992).

20. Jin, H., Oksenberg, D., Ashkenazi, A., Peroutka, S., Duncan, A., Rozmahel, R., Yang, Y., Mengod, G., Palacios, J., and O'Dowd, B. Characterization of the human 5-hydroxytryptamine_{1B} receptor. *J. Biol. Chem.* **267**, 5735-5738 (1992).
21. Marsters, A., Frutkin, A., Simpson, N., Fendly, B. and Ashkenazi, A. Identification of cysteine-rich domains of the type 1 tumor necrosis receptor involved in ligand binding. *J. Biol. Chem.* **267**, 5747-5750 (1992).
22. Chamow, S., Kogan, T., Peers, D., Hastings, R., Byrn, R., and Ashkenazi, A. Conjugation of sCD4 without loss of biological activity via a novel carbohydrate-directed cross-linking reagent. *J. Biol. Chem.* **267**, 15916-15922 (1992).
23. Oksenberg, D., Marsters, A., O'Dowd, B., Jin, H., Havlik, S., Peroutka, S., and Ashkenazi, A. A single amino-acid difference confers major pharmacologic variation between human and rodent 5-HT_{1B} receptors. *Nature* **360**, 161-163 (1992).
24. Haak-Frendscho, M., Marsters, S., Chamow, S., Peers, D., Simpson, N., and Ashkenazi, A. Inhibition of interferon γ by an interferon γ receptor immunoadhesin. *Immunology* **79**, 594-599 (1993).
25. Penica, D., Lam, V., Weber, R., Kohr, W., Basa, L., Spellman, M., Ashkenazi, A., Shire, S., and Goeddel, D. Biochemical characterization of the extracellular domain of the 75-kd tumor necrosis factor receptor. *Biochemistry* **32**, 3131-3138. (1993).
26. Barfod, L., Zheng, Y., Kuang, W., Hart, M., Evans, T., Cerione, R., and Ashkenazi, A. Cloning and expression of a human CDC42 GTPase Activating Protein reveals a functional SH3-binding domain. *J. Biol. Chem.* **268**, 26059-26062 (1993).
27. Chamow, S., Zhang, D., Tan, X., Mhtre, S., Marsters, S., Peers, D., Byrn, R., Ashkenazi, A., and Yunghans, R. A humanized bispecific immunoadhesin-antibody that retargets CD3+ effectors to kill HIV-1-infected cells. *J. Immunol.* **153**, 4268-4280 (1994).
28. Means, R., Krantz, S., Luna, J., Marsters, S., and Ashkenazi, A. Inhibition of murine erythroid colony formation in vitro by iterferon γ and correction by interferon γ receptor immunoadhesin. *Blood* **83**, 911-915 (1994).
29. Haak-Frendscho, M., Marsters, S., Mordenti, J., Gillet, N., Chen, S., and Ashkenazi, A. Inhibition of TNF by a TNF receptor immunoadhesin: comparison with an anti-TNF mAb. *J. Immunol.* **152**, 1347-1353 (1994).

30. Chamow, S., Kogan, T., Venuti, M., Gadek, T., Peers, D., Mordenti, J., Shak, S., and Ashkenazi, A. Modification of CD4 immunoadhesin with monomethoxy-PEG aldehyde via reductive alkylation. *Bioconj. Chem.* **5**, 133-140 (1994).
31. Jin, H., Yang, R., Marsters, S., Bunting, S., Wurm, F., Chamow, S., and Ashkenazi, A. Protection against rat endotoxic shock by p55 tumor necrosis factor (TNF) receptor immunoadhesin: comparison to anti-TNF monoclonal antibody. *J. Infect. Diseases* **170**, 1323-1326 (1994).
32. Beck, J., Marsters, S., Harris, R., Ashkenazi, A., and Chamow, S. Generation of soluble interleukin-1 receptor from an immunoadhesin by specific cleavage. *Mol. Immunol.* **31**, 1335-1344 (1994).
33. Pitti, B., Marsters, M., Haak-Frendscho, M., Osaka, G., Mordenti, J., Chamow, S., and Ashkenazi, A. Molecular and biological properties of an interleukin-1 receptor immunoadhesin. *Mol. Immunol.* **31**, 1345-1351 (1994).
34. Oksenberg, D., Havlik, S., Peroutka, S., and Ashkenazi, A. The third intracellular loop of the 5-HT₂ receptor specifies effector coupling. *J. Neurochem.* **64**, 1440-1447 (1995).
35. Bach, E., Szabo, S., Dighe, A., Ashkenazi, A., Aguet, M., Murphy, K., and Schreiber, R. Ligand-induced autoregulation of IFN- γ receptor β chain expression in T helper cell subsets. *Science* **270**, 1215-1218 (1995).
36. Jin, H., Yang, R., Marsters, S., Ashkenazi, A., Bunting, S., Marra, M., Scott, R., and Baker, J. Protection against endotoxic shock by bactericidal/permeability-increasing protein in rats. *J. Clin. Invest.* **95**, 1947-1952 (1995).
37. Marsters, S., Penica, D., Bach, E., Schreiber, R., and Ashkenazi, A. Interferon γ signals via a high-affinity multisubunit receptor complex that contains two types of polypeptide chain. *Proc. Natl. Acad. Sci. USA* **92**, 5401-5405 (1995).
38. Van Zee, K., Moldawer, L., Oldenburg, H., Thompson, W., Stackpole, S., Montegut, W., Rogy, M., Meschter, C., Gallati, H., Schiller, C., Richter, W., Loetcher, H., Ashkenazi, A., Chamow, S., Wurm, F., Calvano, S., Lowry, S., and Lesslauer, W. Protection against lethal *E. coli* bacteremia in baboons by pretreatment with a 55-kDa TNF receptor-Ig fusion protein, Ro45-2081. *J. Immunol.* **156**, 2221-2230 (1996).
39. Pitti, R., Marsters, S., Ruppert, S., Donahue, C., Moore, A., and Ashkenazi, A. Induction of apoptosis by Apo-2 Ligand, a new member of the tumor necrosis factor cytokine family. *J. Biol. Chem.* **271**, 12687-12690 (1996).

40. Marsters, S., Pitti, R., Donahue, C., Rupert, S., Bauer, K., and Ashkenazi, A. Activation of apoptosis by Apo-2 ligand is independent of FADD but blocked by CrmA. *Curr. Biol.* **6**, 1669-1676 (1996).
41. Marsters, S., Skubatch, M., Gray, C., and Ashkenazi, A. Herpesvirus entry mediator, a novel member of the tumor necrosis factor receptor family, activates the NF- κ B and AP-1 transcription factors. *J. Biol. Chem.* **272**, 14029-14032 (1997).
42. Sheridan, J., Marsters, S., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Ramakrishnan, L., Gray, C., Baker, K., Wood, W.I., Goddard, A., Godowski, P., and Ashkenazi, A. Control of TRAIL-induced apoptosis by a family of signaling and decoy receptors. *Science* **277**, 818-821 (1997).
43. Marsters, S., Sheridan, J., Pitti, R., Gurney, A., Skubatch, M., Baldwin, D., Huang, A., Yuan, J., Goddard, A., Godowski, P., and Ashkenazi, A. A novel receptor for Apo2L/TRAIL contains a truncated death domain. *Curr. Biol.* **7**, 1003-1006 (1997).
44. Marsters, A., Sheridan, J., Pitti, R., Brush, J., Goddard, A., and Ashkenazi, A. Identification of a ligand for the death-domain-containing receptor Apo3. *Curr. Biol.* **8**, 525-528 (1998).
45. Rieger, J., Naumann, U., Glaser, T., Ashkenazi, A., and Weller, M. Apo2 ligand: a novel weapon against malignant glioma? *FEBS Lett.* **427**, 124-128 (1998).
46. Pender, S., Fell, J., Chamow, S., Ashkenazi, A., and MacDonald, T. A p55 TNF receptor immunoadhesin prevents T cell mediated intestinal injury by inhibiting matrix metalloproteinase production. *J. Immunol.* **160**, 4098-4103 (1998).
47. Pitti, R., Marsters, S., Lawrence, D., Roy, Kischkel, F., M., Dowd, P., Huang, A., Donahue, C., Sherwood, S., Baldwin, D., Godowski, P., Wood, W., Gurney, A., Hillan, K., Cohen, R., Goddard, A., Botstein, D., and Ashkenazi, A. Genomic amplification of a decoy receptor for Fas ligand in lung and colon cancer. *Nature* **396**, 699-703 (1998).
48. Mori, S., Marakami-Mori, K., Nakamura, S., Ashkenazi, A., and Bonavida, B. Sensitization of AIDS Kaposi's sarcoma cells to Apo-2 ligand-induced apoptosis by actinomycin D. *J. Immunol.* **162**, 5616-5623 (1999).
49. Gurney, A., Marsters, S., Huang, A., Pitti, R., Mark, M., Baldwin, D., Gray, A., Dowd, P., Brush, J., Heldens, S., Schow, P., Goddard, A., Wood, W., Baker, K., Godowski, P., and Ashkenazi, A. Identification of a new member of the tumor necrosis factor family and its receptor, a human ortholog of mouse GITR. *Curr. Biol.* **9**, 215-218 (1999).

50. Ashkenazi, A., Pai, R., Fong, S., Leung, S., Lawrence, D., Marsters, S., Blackie, C., Chang, L., McMurtrey, A., Hebert, A., DeForge, L., Khoumenis, I., Lewis, D., Harris, L., Bussiere, J., Koeppen, H., Shahrokh, Z., and Schwall, R. Safety and anti-tumor activity of recombinant soluble Apo2 ligand. *J. Clin. Invest.* **104**, 155-162 (1999).
51. Chuntharapai, A., Gibbs, V., Lu, J., Ow, A., Marsters, S., Ashkenazi, A., De Vos, A., Kim, K.J. Determination of residues involved in ligand binding and signal transmission in the human IFN- α receptor 2. *J. Immunol.* **163**, 766-773 (1999).
52. Johnsen, A.-C., Haux, J., Steinkjer, B., Nonstad, U., Egeberg, K., Sundan, A., Ashkenazi, A., and Espevik, T. Regulation of Apo2L/TRAIL expression in NK cells – involvement in NK cell-mediated cytotoxicity. *Cytokine* **11**, 664-672 (1999).
53. Roth, W., Isenmann, S., Naumann, U., Kugler, S., Bahr, M., Dichgans, J., Ashkenazi, A., and Weller, M. Eradication of intracranial human malignant glioma xenografts by Apo2L/TRAIL. *Biochem. Biophys. Res. Commun.* **265**, 479-483 (1999).
54. Hymowitz, S.G., Christinger, H.W., Fuh, G., Ultsch, M., O'Connell, M., Kelley, R.F., Ashkenazi, A. and de Vos, A.M. Triggering Cell Death: The Crystal Structure of Apo2L/TRAIL in a Complex with Death Receptor 5. *Molec. Cell* **4**, 563-571 (1999).
55. Hymowitz, S.G., O'Connell, M.P., Utsch, M.H., Hurst, A., Totpal, K., Ashkenazi, A., de Vos, A.M., Kelley, R.F. A unique zinc-binding site revealed by a high-resolution X-ray structure of homotrimeric Apo2L/TRAIL. *Biochemistry* **39**, 633-640 (2000).
56. Zhou, Q., Fukushima, P., DeGraff, W., Mitchell, J.B., Stetler-Stevenson, M., Ashkenazi, A., and Steeg, P.S. Radiation and the Apo2L/TRAIL apoptotic pathway preferentially inhibit the colonization of premalignant human breast cancer cells overexpressing cyclin D1. *Cancer Res.* **60**, 2611-2615 (2000).
57. Kischkel, F.C., Lawrence, D. A., Chuntharapai, A., Schow, P., Kim, J., and Ashkenazi, A. Apo2L/TRAIL-dependent recruitment of endogenous FADD and Caspase-8 to death receptors 4 and 5. *Immunity* **12**, 611-620 (2000).
58. Yan, M., Marsters, S.A., Grewal, I.S., Wang, H., *Ashkenazi, A., and *Dixit, V.M. Identification of a receptor for BlyS demonstrates a crucial role in humoral immunity. *Nature Immunol.* **1**, 37-41 (2000).

59. Marsters, S.A., Yan, M., Pitti, R.M., Haas, P.E., Dixit, V.M., and Ashkenazi, A. Interaction of the TNF homologues BlyS and APRIL with the TNF receptor homologues BCMA and TACI. *Curr. Biol.* 10, 785-788 (2000).
60. Kischkel, F.C., and Ashkenazi, A. Combining enhanced metabolic labeling with immunoblotting to detect interactions of endogenous cellular proteins. *Biotechniques* 29, 506-512 (2000).
61. Lawrence, D., Shahrokh, Z., Marsters, S., Achilles, K., Shih, D., Mounho, B., Hillan, K., Totpal, K., DeForge, L., Schow, P., Hooley, J., Sherwood, S., Pai, R., Leung, S., Khan, L., Gliniak, B., Bussiere, J., Smith, C., Strom, S., Kelley, S., Fox, J., Thomas, D., and Ashkenazi, A. Differential hepatocyte toxicity of recombinant Apo2L/TRAIL versions. *Nature Med.* 7, 383-385 (2001).
62. Chuntharapai, A., Dodge, K., Grimmer, K., Schroeder, K., Marsters, S.A., Koeppen, H., Ashkenazi, A., and Kim, K.J. Isotype-dependent inhibition of tumor growth in vivo by monoclonal antibodies to death receptor 4. *J. Immunol.* 166, 4891-4898 (2001).
63. Pollack, I.F., Erff, M., and Ashkenazi, A. Direct stimulation of apoptotic signaling by soluble Apo2L/tumor necrosis factor-related apoptosis-inducing ligand leads to selective killing of glioma cells. *Clin. Cancer Res.* 7, 1362-1369 (2001).
64. Wang, H., Marsters, S.A., Baker, T., Chan, B., Lee, W.P., Fu, L., Tumas, D., Yan, M., Dixit, V.M., *Ashkenazi, A., and *Grewal, I.S. TACI-ligand interactions are required for T cell activation and collagen-induced arthritis in mice. *Nature Immunol.* 2, 632-637 (2001).
65. Kischkel, F.C., Lawrence, D.A., Tinel, A., Virmani, A., Schow, P., Gazdar, A., Blenis, J., Arnott, D., and Ashkenazi, A. Death receptor recruitment of endogenous caspase-10 and apoptosis initiation in the absence of caspase-8. *J. Biol. Chem.* 276, 46639-46646 (2001).
66. LeBlanc, H., Lawrence, D.A., Varfolomeev, E., Totpal, K., Morlan, J., Schow, P., Fong, S., Schwall, R., Sinicropi, D., and Ashkenazi, A. Tumor cell resistance to death receptor induced apoptosis through mutational inactivation of the proapoptotic Bcl-2 homolog Bax. *Nature Med.* 8, 274-281 (2002).
67. Miller, K., Meng, G., Liu, J., Hurst, A., Hsei, V., Wong, W-L., Ekert, R., Lawrence, D., Sherwood, S., DeForge, L., Gaudreault, G., Keller, G., Sliwkowski, M., Ashkenazi, A., and Presta, L. Design, Construction, and analyses of multivalent antibodies. *J. Immunol.* 170, 4854-4861 (2003).

68. Varfolomeev, E., Kischkel, F., Martin, F., Wanh, H., Lawrence, D., Olsson, C., Tom, L., Erickson, S., French, D., Schow, P., Grewal, I. and Ashkenazi, A. Immune system development in APRIL knockout mice. Submitted.

Review articles:

1. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D., J. Functional role of muscarinic acetylcholine receptor subtype diversity. *Cold Spring Harbor Symposium on Quantitative Biology*. **LIII**, 263-272 (1988).
2. Ashkenazi, A., Peralta, E., Winslow, J., Ramachandran, J., and Capon, D. Functional diversity of muscarinic receptor subtypes in cellular signal transduction and growth. *Trends Pharmacol. Sci.* Dec Supplement, 12-21 (1989).
3. Chamow, S., Duliege, A., Ammann, A., Kahn, J., Allen, D., Eichberg, J., Byrn, R., Capon, D., Ward, R., and Ashkenazi, A. CD4 immunoadhesins in anti-HIV therapy: new developments. *Int. J. Cancer* Supplement 7, 69-72 (1992).
4. Ashkenazi, A., Capon, and D. Ward, R. Immunoadhesins. *Int. Rev. Immunol.* **10**, 217-225 (1993).
5. Ashkenazi, A., and Peralta, E. Muscarinic Receptors. In *Handbook of Receptors and Channels*. (S. Peroutka, ed.), CRC Press, Boca Raton, Vol. I, p. 1-27, (1994).
6. Krantz, S. B., Means, R. T., Jr., Lina, J., Marsters, S. A., and Ashkenazi, A. Inhibition of erythroid colony formation in vitro by gamma interferon. In *Molecular Biology of Hematopoiesis* (N. Abraham, R. Shadduck, A. Levine F. Takaku, eds.) Intercept Ltd. Paris, Vol. 3, p. 135-147 (1994).
7. Ashkenazi, A. Cytokine neutralization as a potential therapeutic approach for SIRS and shock. *J. Biotechnology in Healthcare* **1**, 197-206 (1994).
8. Ashkenazi, A., and Chamow, S. M. Immunoadhesins: an alternative to human monoclonal antibodies. *Immunomethods: A companion to Methods in Enzymology* **8**, 104-115 (1995).
9. Chamow, S., and Ashkenazi, A. Immunoadhesins: Principles and Applications. *Trends Biotech.* **14**, 52-60 (1996).
10. Ashkenazi, A., and Chamow, S. M. Immunoadhesins as research tools and therapeutic agents. *Curr. Opin. Immunol.* **9**, 195-200 (1997).
11. Ashkenazi, A., and Dixit, V. Death receptors: signaling and modulation. *Science* **281**, 1305-1308 (1998).
12. Ashkenazi, A., and Dixit, V. Apoptosis control by death and decoy receptors. *Curr. Opin. Cell. Biol.* **11**, 255-260 (1999).

13. Ashkenazi, A. Chapters on Apo2L/TRAIL; DR4, DR5, DcR1, DcR2; and DcR3. Online Cytokine Handbook (www.apnet.com/cytokinereference/).
14. Ashkenazi, A. Targeting death and decoy receptors of the tumor necrosis factor superfamily. *Nature Rev. Cancer* 2, 420-430 (2002).
15. LeBlanc, H. and Ashkenazi, A. Apoptosis signaling by Apo2L/TRAIL. *Cell Death and Differentiation* 10, 66-75 (2003).
16. Almasan, A. and Ashkenazi, A. Apo2L/TRAIL: apoptosis signaling, biology, and potential for cancer therapy. *Cytokine and Growth Factor Reviews* 14, 337-348 (2003).

Book:

Antibody Fusion Proteins (Chamow, S., and Ashkenazi, A., eds., John Wiley and Sons Inc.) (1999).

Talks:

1. Resistance of primary HIV isolates to CD4 is independent of CD4-gp120 binding affinity. UCSD Symposium, HIV Disease: Pathogenesis and Therapy. Greenelefe, FL, March 1991.
2. Use of immuno-hybrids to extend the half-life of receptors. IBC conference on Biopharmaceutical Half-life Extension. New Orleans, LA, June 1992.
3. Results with TNF receptor Immunoconjugates for the Treatment of Sepsis. IBC conference on Endotoxemia and Sepsis. Philadelphia, PA, June-1992.
4. Immunoconjugates: an alternative to human antibodies. IBC conference on Antibody Engineering. San Diego, CA, December 1993.
5. Tumor necrosis factor receptor: a potential therapeutic for human septic shock. American Society for Microbiology Meeting, Atlanta, GA, May 1993.
6. Protective efficacy of TNF receptor immunoconjugate vs anti-TNF monoclonal antibody in a rat model for endotoxic shock. 5th International Congress on TNF. Asilomar, CA, May 1994.
7. Interferon- γ signals via a multisubunit receptor complex that contains two types of polypeptide chain. American Association of Immunologists Conference. San Francisco, CA, July 1995.
8. Immunoconjugates: Principles and Applications. Gordon Research Conference on Drug Delivery in Biology and Medicine. Ventura, CA, February 1996.

9. Apo-2 Ligand, a new member of the TNF family that induces apoptosis in tumor cells. Cambridge Symposium on TNF and Related Cytokines in Treatment of Cancer. Hilton-Head, NC, March 1996.
10. Induction of apoptosis by Apo2 Ligand. American Society for Biochemistry and Molecular Biology, Symposium on Growth Factors and Cytokine Receptors. New Orleans, LA, June, 1996.
11. Apo2 ligand, an extracellular trigger of apoptosis. 2nd Clontech Symposium, Palo Alto, CA, October 1996.
12. Regulation of apoptosis by members of the TNF ligand and receptor families. Stanford University School of Medicine, Palo Alto, CA, December 1996.
13. Apo-3: a novel receptor that regulates cell death and inflammation. 4th International Congress on Immune Consequences of Trauma, Shock, and Sepsis. Munich, Germany, March 1997.
14. New members of the TNF ligand and receptor families that regulate apoptosis, inflammation, and immunity. UCLA School of Medicine, LA, CA, March 1997.
15. Immunoadhesins: an alternative to monoclonal antibodies. 5th World Conference on Bispecific Antibodies. Volendam, Holland, June 1997.
16. Control of Apo2L signaling. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. Cold Spring Harbor, New York. September, 1997.
17. Chairman and speaker, Apoptosis Signaling session. IBC's 4th Annual Conference on Apoptosis. San Diego, CA., October 1997.
18. Control of Apo2L signaling by death and decoy receptors. American Association for the Advancement of Science. Philadelphia, PA, February 1998.
19. Apo2 ligand and its receptors. American Society of Immunologists. San Francisco, CA, April 1998.
20. Death receptors and ligands. 7th International TNF Congress. Cape Cod, MA, May 1998.
21. Apo2L as a potential therapeutic for cancer. UCLA School of Medicine. LA, CA, June 1998.
22. Apo2L as a potential therapeutic for cancer. Gordon Research Conference on Cancer Chemotherapy. New London, NH, July 1998.
23. Control of apoptosis by Apo2L. Endocrine Society Conference, Stevenson, WA, August 1998.
24. Control of apoptosis by Apo2L. International Cytokine Society Conference, Jerusalem, Israel, October 1998.

25. Apoptosis control by death and decoy receptors. American Association for Cancer Research Conference, Whistler, BC, Canada, March 1999.
26. Apoptosis control by death and decoy receptors. American Society for Biochemistry and Molecular Biology Conference, San Francisco, CA, May 1999.
27. Apoptosis control by death and decoy receptors. Gordon Research Conference on Apoptosis, New London, NH, June 1999.
28. Apoptosis control by death and decoy receptors. Arthritis Foundation Research Conference, Alexandria GA, Aug 1999.
29. Safety and anti-tumor activity of recombinant soluble Apo2L/TRAIL. Cold Spring Harbor Laboratory Symposium on Programmed Cell Death. . Cold Spring Harbor, NY, September 1999.
30. The Apo2L/TRAIL system: therapeutic potential. American Association for Cancer Research, Lake Tahoe, NV, Feb 2000.
31. Apoptosis and cancer therapy. Stanford University School of Medicine, Stanford, CA, Mar 2000.
32. Apoptosis and cancer therapy. University of Pennsylvania School of Medicine, Philadelphia, PA, Apr 2000.
33. Apoptosis signaling by Apo2L/TRAIL. International Congress on TNF. Trondheim, Norway, May 2000.
34. The Apo2L/TRAIL system: therapeutic potential. Cap-CURE summit meeting. Santa Monica, CA, June 2000.
35. The Apo2L/TRAIL system: therapeutic potential. MD Anderson Cancer Center. Houston, TX, June 2000.
36. Apoptosis signaling by Apo2L/TRAIL. The Protein Society, 14th Symposium. San Diego, CA, August 2000.
37. Anti-tumor activity of Apo2L/TRAIL. AAPS annual meeting. Indianapolis, IN Aug 2000.
38. Apoptosis signaling and anti-cancer potential of Apo2L/TRAIL. Cancer Research Institute, UC San Francisco, CA, September 2000.
39. Apoptosis signaling by Apo2L/TRAIL. Kenote address, TNF family Minisymposium, NIH. Bethesda, MD, September 2000.
40. Death receptors: signaling and modulation. Keystone symposium on the Molecular basis of cancer. Taos, NM, Jan 2001.
41. Preclinical studies of Apo2L/TRAIL in cancer. Symposium on Targeted therapies in the treatment of lung cancer. Aspen, CO, Jan 2001.

42. Apoptosis signaling by Apo2L/TRAIL. Weizmann Institute of Science, Rehovot, Israel, March 2001.
43. Apo2L/TRAIL: Apoptosis signaling and potential for cancer therapy. Weizmann Institute of Science, Rehovot, Israel, March 2001.
44. Targeting death receptors in cancer with Apo2L/TRAIL. Cell Death and Disease conference, North Falmouth, MA, Jun 2001.
45. Targeting death receptors in cancer with Apo2L/TRAIL. Biotechnology Organization conference, San Diego, CA, Jun 2001.
46. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Gordon Research Conference on Apoptosis, Oxford, UK, July 2001.
47. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Cleveland Clinic Foundation, Cleveland, OH, Oct 2001.
48. Apoptosis signaling by death receptors: overview. International Society for Interferon and Cytokine Research conference, Cleveland, OH, Oct 2001.
49. Apoptosis signaling by death receptors. American Society of Nephrology Conference. San Francisco, CA, Oct 2001.
50. Targeting death receptors in cancer. Apoptosis: commercial opportunities. San Diego, CA, Apr 2002.
51. Apo2L/TRAIL signaling and apoptosis resistance mechanisms. Kimmel Cancer Research Center, Johns Hopkins University, Baltimore MD. May 2002.
52. Apoptosis control by Apo2L/TRAIL. (Keynote Address) University of Alabama Cancer Center Retreat, Birmingham, Ab. October 2002.
53. Apoptosis signaling by Apo2L/TRAIL. (Session co-chair) TNF international conference. San Diego, CA, October 2002.
54. Apoptosis signaling by Apo2L/TRAIL. Swiss Institute for Cancer Research (ISREC). Lausanne, Switzerland. Jan 2003.
55. Apoptosis induction with Apo2L/TRAIL. Conference on New Targets and Innovative Strategies in Cancer Treatment. Monte Carlo. February 2003.
56. Apoptosis signaling by Apo2L/TRAIL. Hermelin Brain Tumor Center Symposium on Apoptosis. Detroit, MI. April 2003.
57. Targeting apoptosis through death receptors. Sixth Annual Conference on Targeted Therapies in the Treatment of Breast Cancer. Kona, Hawaii. July 2003.
58. Targeting apoptosis through death receptors. Second International Conference on Targeted Cancer Therapy. Washington, DC. Aug 2003.

Issued Patents:

1. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,329,028 (Jul 12, 1994).
2. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,605,791 (Feb 25, 1997).
3. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 5,889,155 (Jul 27, 1999).
4. Ashkenazi, A., APO-2 Ligand. US patent 6,030,945 (Feb 29, 2000).
5. Ashkenazi, A., Chuntharapai, A., Kim, J., APO-2 ligand antibodies. US patent 6,046,048 (Apr 4, 2000).
6. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,124,435 (Sep 26, 2000).
7. Ashkenazi, A., Chuntharapai, A., Kim, J., Method for making monoclonal and cross-reactive antibodies. US patent 6,252,050 (Jun 26, 2001).
8. Ashkenazi, A. APO-2 Receptor. US patent 6,342,369 (Jan 29, 2002).
9. Ashkenazi, A. Fong, S., Goddard, A., Gurney, A., Napier, M., Tumas, D., Wood, W. A-33 polypeptides. US patent 6,410,708 (Jun 25, 2002).
10. Ashkenazi, A. APO-3 Receptor. US patent 6,462,176 B1 (Oct 8, 2002).
11. Ashkenazi, A. APO-2LI and APO-3 polypeptide antibodies. US patent 6,469,144 B1 (Oct 22, 2002).
12. Ashkenazi, A., Chamow, S. and Kogan, T. Carbohydrate-directed crosslinking reagents. US patent 6,582,928B1 (Jun 24, 2003).



TECHNICAL UPDATE

FROM YOUR LABORATORY SERVICES PROVIDER

HER-2/neu Breast Cancer Predictive Testing

Julie Sanford Hanna, Ph.D. and Dan Mornin, M.D.

EACH YEAR, OVER 182,000 WOMEN in the United States are diagnosed with breast cancer, and approximately 45,000 die of the disease.¹ Incidence appears to be increasing in the United States at a rate of roughly 2% per year. The reasons for the increase are unclear, but non-genetic risk factors appear to play a large role.²

Five-year survival rates range from approximately 65%-85%, depending on demographic group, with a significant percentage of women experiencing recurrence of their cancer within 10 years of diagnosis. One of the factors most predictive for recurrence once a diagnosis of breast cancer has been made is the number of axillary lymph nodes to which tumor has metastasized. Most node-positive women are given adjuvant therapy, which increases their survival. However, 20%-30% of patients without axillary node involvement also develop recurrent disease, and the difficulty lies in how to identify this high-risk subset of patients. These patients could benefit from increased surveillance, early intervention, and treatment.

Prognostic markers currently used in breast cancer recurrence prediction include tumor size, histological grade, steroid hormone receptor status, DNA ploidy, proliferative index, and cathepsin D status. Expression of growth factor receptors and over-expression of the HER-2/neu oncogene have also been identified as having value regarding treatment regimen and prognosis.

HER-2/neu (also known as c-erbB2) is an oncogene that encodes a transmembrane glycoprotein that is homologous to, but distinct from, the epidermal growth factor receptor. Numerous studies have indicated that high levels of expression of this protein are associated with rapid tumor growth, certain forms of therapy resistance, and shorter disease-free survival. The gene has been shown to be amplified and/or overexpressed in 10%-30% of invasive breast cancers and in 40%-60% of intraductal breast carcinoma.³

There are two distinct FDA-approved methods by which HER-2/neu status can be evaluated: immunohistochemistry (IHC, HercepTest™) and FISH (fluorescent in situ hybridization, PathVysion™ Kit). Both methods can be performed on archived and current specimens. The first method allows visual assessment of the amount of HER-2/neu protein present on the cell membrane. The latter method allows direct quantification of the level of gene amplification present in the tumor, enabling differentiation between low- versus high-amplification. At least one study has demonstrated a difference in

recurrence risk in women younger than 40 years of age for low- versus high-amplified tumors (54.5% compared to 85.7%); this is compared to a recurrence rate of 16.7% for patients with no HER-2/neu gene amplification.⁴ HER-2/neu status may be particularly important to establish in women with small (≤ 1 cm) tumor size.

The choice of methodology for determination of HER-2/neu status depends in part on the clinical setting. FDA approval for the Vysis FISH test was granted based on clinical trials involving 1549 node-positive patients. Patients received one of three different treatments consisting of different doses of cyclophosphamide, Adriamycin, and 5-fluorouracil (CAF). The study showed that patients with amplified HER-2/neu benefited from treatment with higher doses of adriamycin-based therapy, while those with normal HER-2/neu levels did not. The study therefore identified a sub-set of women, who because they did not benefit from more aggressive treatment, did not need to be exposed to the associated side effects. In addition, other evidence indicates that HER-2/neu amplification in node-negative patients can be used as an independent prognostic indicator for early recurrence, recurrent disease at any time and disease-related death.⁵ Demonstration of HER-2/neu gene amplification by FISH has also been shown to be of value in predicting response to chemotherapy in stage-2 breast cancer patients.

Selection of patients for Herceptin® (Trastuzumab) monoclonal antibody therapy, however, is based upon demonstration of HER-2/neu protein overexpression using HercepTest™. Studies using Herceptin® in patients with metastatic breast cancer show an increase in time to disease progression, increased response rate to chemotherapeutic agents and a small increase in overall survival rate. The FISH assays have not yet been approved for this purpose, and studies looking at response to Herceptin® in patients with or without gene amplification status determined by FISH are in progress.

In general, FISH and IHC results correlate well. However, subsets of tumors are found which show discordant results; i.e., protein overexpression without gene amplification or lack of protein overexpression with gene amplification. The clinical significance of such results is unclear. Based on the above considerations, HER-2/neu testing at SHMC/PAML will utilize immunohistochemistry (HercepTest®) as a screen, followed by FISH in IHC-negative cases. Alternatively, either method may be ordered individually depending on the clinical setting or clinician preference.

CPT code information

HER-2/neu via IHC

88342 (including interpretive report)

HER-2/neu via FISH

88271x2 Molecular cytogenetics, DNA probe, each

88274 Molecular cytogenetics, interphase in situ hybridization, analyze 25-99 cells

88291. Cytogenetics and molecular cytogenetics, interpretation and report

Procedural Information

Immunohistochemistry is performed using the FDA-approved DAKO antibody kit, Herceptest®. The DAKO kit contains reagents required to complete a two-step immunohistochemical staining procedure for routinely processed, paraffin-embedded specimens. Following incubation with the primary rabbit antibody to human HER-2/neu protein, the kit employs a ready-to-use dextran-based visualization reagent. This reagent consists of both secondary goat anti-rabbit antibody molecules with horseradish peroxidase molecules linked to a common dextran polymer backbone, thus eliminating the need for sequential application of link antibody and peroxidase conjugated antibody. Enzymatic conversion of the subsequently added chromogen results in formation of visible reaction product at the antigen site. The specimen is then counterstained; a pathologist using light-microscopy interprets results.

FISH analysis at SHMC/PAML is performed using the FDA-approved PathVysion™ HER-2/neu DNA probe kit, produced by Vysis, Inc. Formalin fixed, paraffin-embedded breast tissue is processed using routine histological methods, and then slides are treated to allow hybridization of DNA probes to the nuclei present in the tissue section. The PathVysion™ kit contains two direct-labeled DNA probes, one specific for the alphoid repetitive DNA (CEP 17, spectrum orange) present at the chromosome 17 centromere and the second for the HER-2/neu oncogene located at 17q11.2-12 (spectrum green). Enumeration of the probes allows a ratio of the number of copies of chromosome 17 to the number of copies of HER-2/neu to be obtained; this enables quantification of low versus high amplification levels, and allows an estimate of the percentage of cells with HER-2/neu gene amplification. The clinically relevant distinction is whether the gene amplification is due to increased gene copy number on the two chromosome 17 homologues normally present or an increase in the number of chromosome 17s in the cells. In the majority of cases, ratio equivalents less than 2.0 are indicative of a normal/negative result, ratios of 2.1 and over indicate that amplification is present and to what degree. Interpretation of this data will be performed and reported from the Vysis-certified Cytogenetics laboratory at SHMC.

References

1. Wingo, P.A., Tong, T., Bolden, S., "Cancer Statistics", 1995;45:1-3-31.
2. "Cancer Rates and Risks", 4th ed., National Institutes of Health, National Cancer Institute, 1996, p. 120.
3. Slamon, D.J., Clark, G.M., Song, S.G., Levin, W.J., Ulrich, A., McGuire, W.L., "Human breast Cancer: Correlation of relapse and survival with amplification of the her-2/neu oncogene". Science, 235:177-182, 1987.
4. Xing, W.R., Gilchrist, K.W., Harris, C.P., Samson, W., Meisner, L.F., "FISH detection of HER-2/neu oncogene amplification in early onset breast cancer". Breast Cancer Res. And Treatment 39(2):203-212, 1996.
5. Press, M.F., Bernstein, L., Thomas, P.A., Meisner, L.F., Zhou, J.Y., Ma, Y., Haug, G., Robinson, R.A., Harris, C., El-Naggar, A., Slamon, D.J., Phillips, R.N., Ross, J.S., Wolman, S.R., Flom, K.J., "Her-2/neu gene amplification characterized by fluorescence in situ hybridization: poor prognosis in node-negative breast carcinomas", J. Clinical Oncology 15(8):2894-2904, 1997.

Provided for the clients of

**PATHOLOGY ASSOCIATES MEDICAL LABORATORIES
PACLAB NETWORK LABORATORIES
TRI-CITIES LABORATORY
TREASURE VALLEY LABORATORY**

*For more information, please contact
your local representative.*



US006737522B2

(12) **United States Patent**
Sundick et al.

(10) Patent No.: **US 6,737,522 B2**
(45) Date of Patent: ***May 18, 2004**

(54) **CHICKEN INTERLEUKIN-15 AND USES THEREOF**

(75) Inventors: **Roy S. Sundick**, Farmington Hills, MI (US); **Lily A. Jones**, Grosse Point Park, MI (US); **David I. Smith**, Rochester, MN (US)

(73) Assignee: **Wayne State University**, Detroit, MI (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) Appl. No.: **09/907,539**

(22) Filed: **Jul. 17, 2001**

(65) **Prior Publication Data**

US 2002/0150554 A1 Oct. 17, 2002

Related U.S. Application Data

(62) Division of application No. 09/368,613, filed on Aug. 4, 1999, now Pat. No. 6,287,554, and a division of application No. 08/729,004, filed on Oct. 10, 1996, now Pat. No. 6,190,901.

(60) Provisional application No. 60/005,682, filed on Oct. 17, 1995.

(51) Int. Cl.⁷ **C07H 21/04; C12N 15/00; C12N 15/63**

(52) U.S. Cl. **536/23.51; 536/23.1; 435/320.1**

(58) Field of Search **536/23.11, 23.51; 435/320.11, 455.1, 471**

(56) **References Cited**

U.S. PATENT DOCUMENTS

4,490,289 A 12/1984 Stern
4,992,367 A 2/1991 Cullen
5,106,617 A 4/1992 Federicksen et al.
5,149,788 A 9/1992 Bailon et al.
5,162,503 A 11/1992 Bailon et al.
5,455,337 A 10/1995 Devos et al.
5,536,657 A 7/1996 Chua et al.
5,597,901 A 1/1997 Stern

FOREIGN PATENT DOCUMENTS

EP 0 293 180 11/1988 C12P/21/02

EP 0 473 696 B1 7/1995
WO WO95/27722 10/1995
WO WO96/04306 2/1996
WO WO96/37223 2/1996
ZA 9 402 636 11/1994

OTHER PUBLICATIONS

IL-15 Stimulates The Expansion of Aids Virus-specific CTL¹ (The Journal of Immunology, 1996, 157: 3681-3687. Studies Evaluating The Antitumor Activity And Toxicity of Interleukin-15, a New T Cell Growth Factor: Comparison With Interleukin-2 (Immunex Research And Development Corporation, Seattle, Washington 98101 And Oncology And Immunology Research Section, Lederle Laboratories, Pearl River, New York 10965, 1995.

Dialog(R)File 149:iac(sm) Health & Wellness Db(sm) (C) 1997 Info Access Co. (IL-15 Boosts Systemic Immunity Against B7-1 Gene Transfected Breast Cancer Cells (Interleukin 15).

Lillehoj et al., "Avian Interferon and Interleukin-2. A Review by Comparison With Mammalian Homologues" *Poultry Science Research*, vol. 4, No. 2, pp. 67-85, 1992. Interleukin-15, A New T Cell Growth Factor: Comparison with Interleukin-2 (Immunex Research and Development Corporation, Seattle, Washington 98101 and Oncology and Immunology Research Section, Lederle Laboratories, Pearl River, New York 10965) 1995.

Sin et al., J. Virol. vol. 73, 501-509, 1999.

Rodolfo and Colombo, Methods, vol. 19, 114-120, 1999.

Openshaw and Hussell, Dev. Biol. Stand., vol. 92, 179-185, 1998.

Armitage et al., "IL-15 Has Stimulatory Activity for the Induction of B Cell Proliferation and Differentiation" *Journal of Immunology*, vol. 154, No. 2, pp. 483-490, 1995.

Gini et al., "IL-15 A Novel T Cell Growth Factor That Shares Activities and Receptor Components With IL-2" *Journal of Leukocyte Biology*, vol. 57, No. 5, pp. 763-766, 1995.

Staeheli et al., "Cytokines of Birds: Conserved Functions-A Largely Different Look" *Journal of Interferon and Cytokine*, vol. 21, No. 12, pp. 993-1010, 2001.

Primary Examiner—David Guzo

Assistant Examiner—Quang Nguyen

(74) *Attorney, Agent, or Firm*—Darby & Darby

(57) **ABSTRACT**

The present invention pertains to isolated DNA encoding avian interleukin-15 and to purified interleukin-15 polypeptides.

27 Claims, 3 Drawing Sheets

FIG. 1A

T7 end of the pCDNA1 vector:

5' -TGCTTGGTACCGAGCTCGGATCCACTAGTAACGCCCGCCAGTGTGCTCTAAAG-

Noncoding Segment of cDNA: *CAGATAACTGGGACACTGCC

Coding Region of First Open Region Frame (IL-15):

ATGATGTGCAAAAGTACTGATCTTTGGCTGTATTTTCGGTAGCAACGCTAATG

ACTACAGCTTATGGAGCATCTCTATCATCAGCAAAAAGGAAACCTCTTCAA

ACATTAAATAAGGATTAGAAATATTGGAATAATATCAAGAACAAAGATTTCAT

CTCGAGCTCTACACACCAACTGAGACCCAGGAGTGCACCCAGCAAACTCTG

CAGTGTTACCTGGGAGAAGTGGTTACTCTGAAGAAAGAAACTGAAGATGAC

ACTGAAATTAAAGAAGAATTGTAACTGCTATTCAAATAATATCGAAAAGAAC

CTCAAGAGTCTTACGGGTCTAAATCACACCCGGAAGTGAATGCAAGATCTGT

GAAGCTAACACAAGAAAAAATTTCCTGATTTTCTCCATGAAGTGAACCAAC

TTTGTGAGATATCTGCAAAA

A _____ A

FIG. 1B

A ————— A

Sequence of Remaining Insert cDNA:

TAAGCAACTAATCATTTTATTATTTTACTGCTATGTTATTTATTTAATTATTT
AATTACAGATAAATTATATATTTTATCCCGTGGCTAACTAATCTGCTGTCC
ATTCTGGGACCACTGTATGCTCTTAGTCTGGGTGATAGACGCTGTCTTA
AGATCATATTTGATCCTTTCTGTAACTACGGGCTCAAAATGTACGTTGGA
AAACTGATTGATTCTCACTTTGTGCGGTAAAGTGATATGTGTTTACTGAAAG
AATTTTAAAGTCACTTCTAGATGACATTTAATAAATTTCAG#

Sp6 end of the pCDNA1 vector:

CTTAGACACACTGGCGGCCNTCGAGCATGCATCTAGAGGGCC-3'

* beginning of cDNA

end of cDNA

FIG. 2

chicken IL-15 precursor, 143 amino acids

MMCKVLI FGCISVATLMTTAYGASLSSAKRKPLQTLIKDLEILENIKNKI
HLELYTPTETQECTQQTLQCYLGEVVTLLKKETEDDTEIKEEFVTAIQNIE
KNLKS LTGLNHTGSECKICEANNKKFPDFLHEL TNFVRYLQK

CHICKEN INTERLEUKIN-15 AND USES THEREOF

This is a divisional of U.S. patent application Ser. No. 09/368,613 filed Aug. 4, 1999, which issued as U.S. Pat. No. 6,287,554 and is a divisional of U.S. patent application Ser. No. 08/729,004, filed on Oct. 10, 1996 and now issued as U.S. Pat. No. 6,190,901. U.S. patent application Ser. No. 08/729,004 claims priority to U.S. provisional patent application Serial No. 60/005,682 filed on Oct. 17, 1995. Each of these prior applications is hereby incorporated herein by reference, in its entirety.

FIELD OF INVENTION

The present invention pertains to isolated genes encoding avian interleukin-15 and to purified interleukin-15 polypeptides.

BACKGROUND OF THE INVENTION

Most chickens produced in developed countries for consumption and egg-laying (at least 10 billion per year) are vaccinated to protect them against Marek's disease. All of the egg-laying chickens and breeder stocks are also vaccinated with Newcastle Disease Virus, Infectious Bursal Disease Virus, Infectious Bronchitis Virus, Fowlpox Virus and Coccidial vaccines. For optimal protection, Marek's vaccination is performed either at or before hatching. One obstacle to the development of efficacious pre-hatching and at-hatching vaccination regimens is that the embryonic and newly hatched avian immune system is not fully developed and cannot mount as effective an immune response to the immunogen as at 2-3 weeks after hatching. Thus, there is a need in the art for agents and compositions that enhance the effectiveness of pre- and post-hatching avian vaccines.

Interleukin-2 and interleukin-15 are related cytokines that stimulate the activity and proliferation of T cells in mammals. Though IL-2 and IL-15 both interact with the β and γ chains of the IL-2 receptor, and may share some elements of tertiary structure, the two polypeptides are not homologous and represent distinct gene products.

The genes encoding IL-15 from several different mammalian species share a high degree of homology. For example, human and simian IL-15 share 97% amino acid homology. By contrast, chicken IL-15, which is the subject of the present invention, shares only 25% amino acid identity with mammalian IL-15. Another distinguishing characteristic of chicken IL-15 is that it (and not the mammalian forms) is produced by mitogen-activated spleen cells. Accordingly, the discovery of chicken IL-15 and the finding that it possesses T cell-stimulatory activity provide a novel reagent for vaccine augmentation in avian species. Without wishing to be bound by theory, the bioactivity of mammalian IL-15 in stimulating skeletal muscle development suggests that avian IL-15s are also useful in stimulating growth in avian species.

SUMMARY OF THE INVENTION

The present invention provides isolated and purified DNA encoding avian interleukin-15 (IL-15), as well as cloning and expression vectors comprising IL-15 DNA and cells transformed with IL-15-encoding vectors. Avian species from which IL-15 may be derived include without limitation chicken, turkey, duck, goose, quail and pheasant.

The invention also provides isolated and purified avian IL-15 polypeptide, the native secreted or mature form of

which has a molecular mass of about 14 kDa, an isoelectric point of about 6.57, a net charge of -2, and a hydrophilicity index of 0.278, and which has the ability to stimulate mitogen-activated avian T cells and to promote the growth of other cell types. IL-15 according to the present invention may be obtained from native or recombinant sources.

Also encompassed by the invention are sequence-conservative and function-conservative variants of avian IL-15 DNA and IL-15 polypeptides, including, for example, a bioactive IL-15 sequence or sub-fragment that is fused in-frame to a purification sequence.

In another aspect, the invention provides a method for enhancing an immune response in fowl to an immunogen, which is achieved by administering the immunogen before, after, or substantially simultaneously with avian IL-15 in an amount effective to enhance the immune response.

In yet another aspect, the invention provides a vaccine for inducing an immune response in fowl to an immunogen, comprising the immunogen and an effective amount of avian interleukin-15 for immune response enhancement. The immunogen may be derived, for example, from avian pathogens such as Marek's Disease Virus, Newcastle Disease Virus, Infectious Bursal Disease Virus, Infectious Bronchitis Virus, and the like.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is an illustration of an 845 nt sequence including 747 nt of cDNA sequence encoding chicken interleukin-15 (IL-15) SEQ ID NO:1.

FIG. 2 is an illustration of a 143-amino acid sequence corresponding to the chicken interleukin-15 precursor polypeptide (SEQ ID NO:2).

DETAILED DESCRIPTION OF THE INVENTION

All patent applications, patents, and literature references cited in this specification are hereby incorporated by reference in their entirety. In case of conflict, the present description, including definitions, will control.

The present invention encompasses interleukin-15 (IL-15) from avian species. The invention provides isolated and purified nucleic acids encoding avian IL-15, as well as IL-15 polypeptides purified from either native or recombinant sources. Avian IL-15 produced according to the present invention may be used in commercial fowl cultivation to promote growth and to enhance the efficacy of avian vaccines.

Nucleic Acids, Vectors, Transformants

The sequence of the cDNA encoding chicken IL-15 is shown in FIG. 1 (SEQ ID NO:1), and the predicted amino acid sequence of chicken IL-15 is shown in FIG. 2 (SEQ ID NO:2). The designation of this avian polypeptide as IL-15 is based on partial amino acid sequence homology to mammalian IL-15 and the ability of the polypeptide to stimulate mitogen-activated T cells (see below). Furthermore, without wishing to be bound by theory, it is predicted that avian IL-15 polypeptides also exhibit one or more of the following bioactivities: activation of NK (natural killer) cells, stimulation of B-Cell maturation, proliferation of mast cells, and interaction with the beta and gamma subunits of the IL-2 receptor.

Because of the degeneracy of the genetic code (i.e., multiple codons encode certain amino acids), DNA sequences other than that shown in FIG. 1 can also encode the chicken IL-15 amino acid sequences shown in FIG. 2.

Such other DNAs include those containing "sequence-conservative" variations in which a change in one or more nucleotides in a given codon results in no alteration in the amino acid encoded at that position. Furthermore, a given amino acid residue in a polypeptide can often be changed without altering the overall conformation and function of the native polypeptide. Such "function-conservative" variants include, but are not limited to, replacement of an amino acid with one having similar physico-chemical properties, such as, for example, acidic, basic, hydrophobic, and the like (e.g., replacement of lysine with arginine, aspartate with glutamate, or glycine with alanine). In addition, amino acid sequences may be added or deleted without destroying the bioactivity of the molecule. For example, additional amino acid sequences may be added at either amino- or carboxy-terminal ends to serve as purification tags, (i.e., to allow one-step purification of the protein, after which they may be chemically or enzymatically removed). Alternatively, the additional sequences may confer an additional cell-surface binding site or otherwise alter the target cell specificity of IL-15.

The chicken IL-15 cDNAs within the scope of the present invention are those of FIG. 1, sequence-conservative variant DNAs, DNA sequences encoding function-conservative variant polypeptides, and combinations thereof. The invention encompasses fragments of avian interleukin-15 that exhibit a useful degree of bioactivity, either alone or in combination with other sequences or components. As explained below, it is well within the ordinary skill in the art to predictively manipulate the sequence of IL-15 and establish whether a given avian IL-15 variant possesses an appropriate stability and bioactivity for a given application. This can be achieved by expressing and purifying the variant IL-15 polypeptide in a recombinant system and assaying its T-cell stimulatory activity and/or growth-promoting activity in cell culture and in animals, followed by testing in the application.

The present invention also encompasses IL-15 DNAs (and polypeptides) derived from other avian species, including without limitation ducks, turkeys, pheasants, quail and geese. Avian IL-15 homologues of the chicken sequence shown in FIG. 1 are easily identified by screening cDNA or genomic libraries to identify clones that hybridize to probes comprising all or part of the sequence of FIG. 1. Alternatively, expression libraries may be screened using antibodies that recognize chicken IL-15. Without wishing to be bound by theory, it is anticipated that IL-15 genes from other avian species will share at least about 70% homology with the chicken IL-15 gene. Also within the scope of the invention are DNAs that encode chicken homologues of IL-15, defined as DNA encoding polypeptides that share at least about 25% amino acid identity with chicken IL-15.

Generally, nucleic acid manipulations according to the present invention use methods that are well known in the art, such as those as disclosed in, for example, *Molecular Cloning, A Laboratory Manual* (2nd Ed., Sambrook, Fritsch and Maniatis, Cold Spring Harbor), or *Current Protocols in Molecular Biology* (Eds. Ausubel, Brent, Kingston, More, Feidman, Smith and Stuhl, Greene Publ. Assoc., Wiley-Interscience, NY, N.Y., 1992).

The present invention encompasses cDNA and RNA sequences and sense and antisense sequences. The invention also encompasses genomic avian IL-15 polypeptide DNA sequences and flanking sequences, including, but not limited to, regulatory sequences. Nucleic acid sequences encoding avian IL-15 polypeptide(s) may also be associated with heterologous sequences, including promoters, enhancers,

response elements, signal sequences, polyadenylation sequences, introns, 5'- and 3'- noncoding regions, and the like. Transcriptional regulatory elements that may be operably linked to avian IL-15 polypeptide DNA sequence(s) include without limitation those that have the ability to direct the expression of genes derived from prokaryotic cells, eukaryotic cells, viruses of prokaryotic cells, viruses of eukaryotic cells, and any combination thereof. Other useful heterologous sequences are known to those skilled in the art.

The nucleic acids of the present invention can be modified by methods known to those skilled in the art to alter their stability, solubility, binding affinity, and specificity. For example, the sequences can be selectively methylated. The nucleic acid sequences of the present invention may also be modified with a label capable of providing a detectable signal, either directly or indirectly. Exemplary labels include radioisotopes, fluorescent molecules, biotin, and the like.

The present invention also provides vectors that include nucleic acids encoding the avian IL-15 polypeptide(s). Such vectors include, for example, plasmid vectors for expression in a variety of eukaryotic and prokaryotic hosts. Preferably, vectors also include a promoter operably linked to the avian IL-15 polypeptide encoding portion. The encoded avian IL-15 polypeptide(s) may be expressed by using any suitable vectors and host cells as explained herein or otherwise known to those skilled in the art.

Vectors will often include one or more replication systems for cloning or expression, one or more markers for selection in the host such as, for example, antibiotic resistance, and one or more expression cassettes. The inserted coding sequences may be synthesized, isolated from natural sources, prepared as hybrids, or the like. Ligation of the coding sequences to the transcriptional regulatory sequences may be achieved by methods known to those skilled in the art. Suitable host cells may be transformed/transfected/infected by any suitable method including electroporation, CaCl_2 - or liposome-mediated DNA uptake, fungal infection, microinjection, microprojectile, or the like.

Suitable vectors for use in practicing the present invention include without limitation YEpl352, pCDNA1 (In Vitrogen, San Diego, Calif.), pRc/CMV (In Vitrogen), and pSFV1 (GIBCO/BRL, Gaithersburg, Md.). One preferred vector for use in the invention is pSFV1. Suitable host cells include *E. coli*, yeast, COS cells, PC12 cells, CHO cells, GH4C1 cells, BHK-21 cells, and amphibian melanophore cells. BHK-21 cells are a preferred host cell line for use in practicing the present invention.

Nucleic acids encoding avian IL-15 polypeptide(s) may also be introduced into cells by recombination events. For example, such a sequence can be microinjected into a cell, effecting homologous recombination at the site of an endogenous gene encoding the polypeptide, an analog or pseudogene thereof, or a sequence with substantial identity to an avian IL-15 polypeptide-encoding gene. Other recombination-based methods such as non-homologous recombinations, and deletion of endogenous gene by homologous recombination, especially in pluripotent cells, may also be used.

IL-15 Polypeptides

The chicken IL-15 gene (the cDNA of which is shown in FIG. 1) encodes a polypeptide of 143 amino acids (FIG. 2). Without wishing to be bound by theory, by comparison with simian IL-15, and by use of an accepted procedure to predict signal peptidase cleavage sites (Von Heijne, *Nuc. Acids Res.*, 14:4683, 1986), it is predicted that an aminoterminal leader sequence of about 22 amino acids (secretion signal peptide)

is cleaved from the primary translation product to produce mature IL-15. The predicted mature sequence of 121 amino acids is further characterized by a predicted molecular weight of 13,971 daltons; an isoelectric point of 6.57; four cysteine residues (at amino acids numbers 63, 70, 116, and 119 in the precursor IL-15 shown in FIG. 2) that correspond to four cysteines conserved among human, mouse, and monkey IL-15 and that are believed to participate in intramolecular disulfide bonding; and one consensus site for N-linked glycosylation (at asparagine 110 of the sequence shown in FIG. 2) which corresponds to a similar site in human IL-15.

Purification of IL-15 from natural or recombinant sources may be achieved by methods well-known in the art, including without limitation ion-exchange chromatography, reverse-phase chromatography on C4 columns, gel filtration, isoelectric focusing, affinity chromatography, immunoaffinity chromatography, and the like. In a preferred embodiment, large quantities of bioactive IL-15 may be obtained by constructing a recombinant DNA sequence comprising the coding region for IL-15 fused in frame to a sequence encoding 6 C-terminal histidine residues in the pSFV1 replicon (GIBCO/BRL). mRNA encoded by this plasmid is synthesized using techniques well-known to those skilled in the art and introduced into BHK-21 cells by electroporation. The cells synthesize and secrete mature glycosylated IL-15 polypeptides containing 6 C-terminal histidines. The modified IL-15 polypeptides are easily purified from the cell supernatant by affinity chromatography using a histidine-binding resin (His-bind, Novagen, Madison, Wis.).

Avian IL-15 polypeptides isolated from any source can be modified by methods known in the art. For example, avian IL-15 may be phosphorylated or dephosphorylated, glycosylated or deglycosylated, and the like. Especially useful are modifications that alter avian IL-15 solubility, stability, and binding specificity and affinity.

Anti-IL-15 Antibodies

The present invention encompasses antibodies that are specific for avian IL-15 polypeptides identified as described above. The antibodies may be polyclonal or monoclonal, and may discriminate avian IL-15s from different species, identify functional domains, and the like. Such antibodies are conveniently made using the methods and compositions disclosed in Harlow and Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988, other references cited herein, as well as immunological and hybridoma technologies known to those in the art. Where natural or synthetic avian IL-15-derived peptides are used to induce an avian IL-15-specific immune response, the peptides may be conveniently coupled to a suitable carrier such as KLH and administered in a suitable adjuvant such as Freund's. Preferably, selected peptides are coupled to a lysine core carrier substantially according to the methods of Tam (1988) *Proc. Natl. Acad. Sci. USA*, 85:5409-5413. The resulting antibodies may be modified to a monovalent form e.g. Fab, FAB', or FV. Anti-idiotypic antibodies, especially internal imaging anti-idiotypic antibodies, may also be prepared using known methods.

In one embodiment, purified avian IL-15 is used to immunize mice, after which their spleens are removed, and splenocytes used to form cell hybrids with myeloma cells to obtain clones of antibody-secreting cells according to techniques that are standard in the art. The resulting monoclonal antibodies secreted by such cells are screened using in vitro assays for the following activities: binding to avian IL-15, inhibiting the receptor-binding activity of IL-15, and inhibiting the T-cell stimulatory activity of IL-15.

Anti-avian IL-15 antibodies may be used to identify and quantify avian IL-15, using immunoassays such as ELISA, RIA, and the like. Anti-avian IL-15 antibodies may also be used to immunodeplete extracts of avian IL-15. In addition, these antibodies can be used to identify, isolate, and purify avian IL-15s from different sources, and to perform subcellular and histochemical localization studies.

Applications

Avian IL-15 produced according to the present invention can be used beneficially in homologous or heterologous avian species, for example, to stimulate activated T-cells (Grabstein et al., *Science*, 264:965, 1994) and B-cells (Armitage et al., *J. Immunol.*, 154:483, 1995) and/or to promote the growth of non-immune cells, such as, for example, muscle cells (Quinn et al., *Endocrinol.* 136:3669, 1995).

Vaccines

The present invention encompasses methods and compositions for enhancing the efficacy of an immune response in avian species. In this embodiment, avian IL-15 is used in conjunction with an immunogen for which it is desired to elicit an immune response. For example, in avian vaccines, such as those against Marek's disease, Newcastle Disease Virus, and other pathogens such as Infectious Bursal Disease Virus and Infectious Bronchitis Virus, it is desirable to include avian IL-15 in the vaccine to enhance the magnitude and quality of the immune response. For this purpose, IL-15 purified from native or recombinant sources as described above is included in the vaccine formulation at a concentration ranging from about 0.01 μ g to about 1.0 μ g per vaccine per chicken.

IL-15 may be administered in conjunction with a live (i.e., replicating) vaccine or a non-replicating vaccine. Non-limiting examples of replicating vaccines are those comprising native or recombinant viruses or bacteria, such as modified turkey herpesvirus or modified fowlpox virus. Non-limiting examples of non-replicating vaccines are those comprising killed or inactivated viruses or other microorganisms, or crude or purified antigens derived from native, recombinant, or synthetic sources, such as, for example, coccidial vaccines. Commercial sources for avian vaccines include without limitation: Rhone Merieux Laboratoire-IFFA (Lyon, France); Intervet International BV (Boxmeer, The Netherlands); Mallinckrodt Veterinary; Solvay Animal Health (Mendota Heights, Minn.); Hoechst-Roussel (Knoxville, Tenn.); and Nippon Zeon Co., Ltd. (Kawasaki-Kiu, Japan).

In one embodiment, the gene encoding IL-15 is incorporated into a recombinant virus, which is then formulated into a live vaccine. The IL-15 gene is incorporated into the virus so that its expression is controlled by an appropriate promoter. Administration of the vaccine results in the expression of bioactive IL-15 in close temporal and spatial proximity to the desired immune response, thus enhancing the vaccine's efficacy.

IL-15 may be administered to birds as part of a vaccine formulation either before or after hatching, preferably before hatching, using methods known in the art such as those described in U.S. Pat. Nos. 5,034,513 and 5,028,421.

Growth Promotion

The present invention provides methods and compositions for enhancing the growth of avian species for medical and/or commercial purposes. In this embodiment, IL-15 is administered to birds using any appropriate mode of administration. For growth promotion, IL-15 is administered in amounts ranging from about 0.25 μ g/kg/day to about 25 μ g/kg/day. It will be understood that the required amount of

IL-15 can be determined by routine experimentation well-known in the art, such as by establishing a matrix of dosages and frequencies and comparing a group of experimental units or subjects to each point in the matrix.

According to the present invention, native or recombinant avian IL-15 may be formulated with a physiologically acceptable carrier, such as, for example, phosphate buffered saline or deionized water. The formulation may also contain excipients, including lubricant(s), plasticizer(s), colorant(s), absorption enhancer(s), bactericide(s), and the like that are well-known in the art. The IL-15 polypeptide of the invention may be administered by any effective means, including without limitation intravenous, subcutaneous, intramuscular, transmucosal, topical, or oral routes. For subcutaneous administration, for example, the dosage form may consist of IL-15 in sterile physiological saline. For oral administration, IL-15, with or without excipients, may be micro- or macro-encapsulated in, e.g., liposomes and microspheres. Dermal patches (or other slow-release dosage forms) may also be used.

The following examples are intended to further illustrate the invention without limiting its scope thereof.

EXAMPLE 1

Cloning of the Chicken IL-15 Gene

To clone chicken IL-15, a chicken spleen cell cDNA library derived from spleen cells that had been activated with concanavalin A was utilized (Kaplan, *J. Immunol.* 151:628, 1993). 5000 colonies were grown overnight at 35° C. on LB agar plates containing 30 µg/ml ampicillin and 10 µg/ml tetracycline. 15–20 colonies were pooled and transferred to 10 ml Terrific Broth (containing the same antibiotics) and grown overnight. Plasmid DNA from each pool was then isolated by published procedures (Maniatis, Section 1.28), treated with RNAase (10 µg/ml), and stored in TE buffer.

The plasmid DNAs were transfected into COS-7(ATCC) cells using Lipofectamine (GIBCO/BRL, Gaithersburg, Md.). 1 µg of each plasmid pool was mixed with 3 µl Lipofectamine in 100 µl Opti-MEM medium (GIBCO/BRL), incubated for 30 min, and then placed on COS-7 cells that had been grown to 80–90% confluence in 12-well plates and rinsed in serum-free medium. The cells and DNA were incubated for 5 hrs at 37° C. with Dulbecco's MEM in the absence of serum and antibiotics, and then supplemented with the same medium containing 10% fetal calf serum and incubated overnight at 37° C. The next day, the medium was replaced with Dulbecco's MEM containing 10% fetal calf serum, penicillin, and streptomycin. After an additional 24 hrs of incubation, the medium was collected and stored at –20° C.

The cell supernatants were tested for IL-15 activity as described in Example 2 below. Five pools with the highest stimulation indices (1.6 to 2.1) exhibited levels of activity that were greater than 2 standard deviations from the mean of the remaining 278 pools. Three of the five pools remained positive in a second screen, and were subdivided into pools of 6. Plasmid DNA extracted from the secondary pools was used to transfect COS-7 cells and the supernatants were tested for IL-2-like activity. As described below in Example 2, three positive pools were identified and subdivided to yield individual clones; from each pool at least one positive clone was isolated.

The complete cDNA inserts of all three positive clones were sequenced using the automated Applied Biosystems

Model 373A sequencing system. The flanking T7 and SP6 primers contained in the pcDNA1 vector were used to prime the sequencing reaction. Two of the clones, B2.16.2 and M2.12.1, were identical and coded for the cDNA sequence shown in FIG. 1. Clone F19.84 was similar to those two clones, but was missing the 20 nt at its 5' end (i.e., starting at the first ATG of the coding region) and contained a poly T tail of at least 100 nt at its 3' end.

The entire 747 nt sequence (FIG. 1, SEQ ID NO:1) was analyzed using a BLAST search (which accesses all of the major international nucleotide data banks). No significant homology was detected with any other known sequence. The sequence was also analyzed using the MacVector software program (MacVector 4.0; International Biotechnologies, Inc., New Haven, Conn.) on a Mac IIci computer. This analysis revealed an open reading frame flanked at its 5' end by a Kozak consensus sequence for translation initiation. The predicted amino acid sequence of this open reading frame is shown in FIG. 2 (Seq ID NO:2). This amino acid sequence was analyzed using a BLASTP search (which accesses all of the major international protein data banks) revealing significant homology with monkey and human precursor IL-15.

The predicted amino acid sequence of chicken IL-15 consists of a 143 amino acid polypeptide having a predicted molecular weight of 16,305 and an isoelectric point of 6.37. Based on the hydrophobicity of its amino terminal end and by comparison with known signal peptide cleavage sites (von Heijne, *Nucleic Acids Res.* 14:4683, 1986) it is predicted that cleavage between glycine-22 and alanine-23 results in the removal of an aminoterminal leader sequence of about 22 amino acids (secretion signal peptide) from the primary product to produce mature IL-15.

The predicted mature IL-15 sequence of 121 amino acids has a predicted molecular weight of 13,971, an isoelectric point of 6.57, and a possible N-linked glycosylation site (at asparagine 110 of FIG. 2). Comparisons between the predicted amino acid sequences of IL-15 from monkey, human, mouse and chicken and analysis of the tertiary structure of monkey IL-15 (Grabstein, *Science*, 264:965, 1994) suggest that four cysteines in chicken IL-15 (positions 63, 70, 116 and 119 of precursor IL-15, FIG. 2) are conserved and form intrachain disulfide bonds.

EXAMPLE 2

Bioactivity Assay for Chicken IL-15

Bioactivity assays for IL-15 are performed as follows: Concanavalin A (ConA)-activated splenic T cells are prepared by incubating chicken spleen cells (10⁷ cells/ml) with Con A (10 µg/ml) (Sigma Chemical Co., St. Louis, Mo.) in RPMI 1640 medium (Sigma) containing 2 mg/ml BSA, antibiotics and glutamine at 40° C. for 24 hrs. The medium is then replaced with Iscoves' medium (Sigma) containing 2% normal chicken serum (Sigma) and 0.05M alpha-methyl pyranoside (Sigma) for an additional 2–4 days, diluting the cells in additional medium as needed. Blast cells are purified from this mixture by gently layering them on a Histopaque density gradient (Sigma) and centrifuging them according to the manufacturer's instructions. The cells are then washed three times and finally resuspended in assay medium (Iscoves' containing 2% normal chicken serum (Sigma)).

For the assay, 2×10⁴ blast cells are placed in roundbottom 96 well plates in assay medium containing IL-15 (such as, e.g., dilutions of supernatant from transfected COS-7 cells) or appropriate controls. After overnight incubation at 40° C.,

9

the cells are pulsed for 6 hrs with ^3H -thymidine (0.5 μCi) (New England Nuclear, Boston, Mass.)+fluorodeoxyuridine (10^{-6}M) (Sigma). The cells are then harvested on glass fiber filters (Whatman, Clifton, N.J.), and the radioactivity is measured in a liquid scintillation counter. IL-15 is expressed as a stimulation index, which is the radioactivity in experimental samples—the radioactivity in controls (non-transfected COS-7 supernatants). A typical result is shown in Table 1.

TABLE 1

SOURCE OF PLASMID DNA		Stimulation indices			
DNA	Designation	1/10 dil ^a	1/10 dil ^b	1/33 dil ^b	1/100 dil ^b
PRIMARY POOLS	A19	1.6	1.9	1.3	1.2
	B2	2.1	4.2	2.3	1.7
	E7	1.8	1.7	1.5	0.9
	F19	1.8	3.5	2.0	1.2
	M2	1.7	3.2	1.9	1.3
	Ave. of 278 \pm SD	1.1 \pm 0.1			
SECONDARY POOLS	Ave. of 3 Neg. pools		1.4	1.3	1.1
	A19.7		0.7	1.9	
	B2.16		6.0	3.5	
	F19.8		9.8	3.4	
	M2.12		3.2	2.2	
INDIVIDUAL CLONES	B2.16.2		6.6	3.3	2.7
	F19.8.4		7.5	4.0	3.0
	M2.12.1		7.2	3.9	3.6

^aFirst screening at 1/10 dil.

^bA repeat transfection using 5 positive and 3 negative primary pools

EXAMPLE 3

Expression and Purification of IL-15

To obtain high-level expression of chicken IL-15 in mammalian cells, the pSFV1 eukaryotic expression vector (which includes the Semliki Forest Virus replicon) is used (GIBCO/BRL, Gaithersburg, Md.). Use of this vector allows for signal peptide cleavage, glycosylation, and secretion of mature active protein. In one embodiment, the recombinant vector encodes an additional six histidine residues at the carboxyterminus of the native IL-15 sequence, allowing the efficient single step purification of the secreted protein on a nickel column (Novagen, Madison, Wis.).

Primers were constructed that include 5' and 3' sequences flanking the coding region of IL-15 cDNA. The 3' primer also includes nucleotides coding for 6 histidines. These primers were used in polymerase chain reaction (PCR), using as a template the entire IL-2 cDNA contained within the pcDNA1 plasmid. The resulting amplified cDNA, including the histidine-coding sequences, was ligated into the pSFV1 plasmid (GIBCO/BRL). The plasmid was obtained by transforming DH5 *E. coli* (GIBCO/BRL) and selecting transformants on agar plates and broth containing ampicillin.

This plasmid is used as a template to produce mRNA in vitro, using manufacturer's protocols. The mRNA is transfected into BHK-21 cells by electroporation, using 10 μg RNA per 10^7 cells, after which the cells are incubated for 1–3 days. The cell supernatant is harvested and passed through a resin matrix (His-Bind resin; Novagen, Madison, Wis.) using a suitable buffer system (His-bind buffer kit; Novagen). Up to 20 mg of tagged protein can be purified on a single 2.5 ml column. The IL-15 is eluted from the column with the elution buffer provided in the kit. It is estimated that BHK-21 cells growing in 50 ml medium synthesize about 25

10

mg total protein, with up to 5% comprising a recombinantly expressed and secreted protein. This corresponds to approximately 1.25 mg of cIL-15.

EXAMPLE 4

Use of Avian IL-15 in Vaccines

The following experiments are performed to evaluate the immune-enhancing activity of chicken IL-15 in chicken vaccines.

Chicken IL-15 cDNA is inserted into two viral vectors (derived from turkey herpesvirus and fowlpox virus, respectively) that are used for the expression of recombinant proteins in chickens (Morgan et al., *Avian Diseases*, 36:858, 1992; Yanagida et al., *J. Virol.*, 66:1402, 1992; Nazerian et al., *J. Virol.*, 66:1409, 1992). These IL-15-modified live viral vectors are administered to newly hatched chicks simultaneously with the administration of various vaccines currently available. Six days later the chicks are challenged with the corresponding virulent viruses and observed for 8 weeks for the development of disease. The incidence of disease in these chicks is compared with controls that do not receive the IL-15-modified live viral vectors. A sample protocol (including expected results) is shown in Table 2.

TABLE 2

Group #	Treatment on day 1	Challenge at day 6	% expected with disease
1	none	none	0
2	none	virulent Marek's	>80%
3	HVT (not modified)	virulent Marek's	20%
4	HVT-IL-15 ^a	virulent Marek's	0 to 10%
5	HVT (not modified) + HVT-IL-15	virulent Marek's	0 to 10%
6	none	virulent NDV	>80%
7	HVT-IF-15	virulent NDV	30% to >50%
8	NDV vaccine	virulent NDV	20%
9	NDV vaccine + HVT-IL-15	virulent NDV	0 to 10%

^aherpesvirus of turkeys expressing IL-15

In an alternative procedure, newly hatched chicks are injected intramuscularly with 100 μg of a plasmid containing cDNA for chicken IL-15, using the methods described in Ulmer, *J. B. Science*, 259:1745–1749, 1993. These chicks, and control chicks receiving a control vector lacking IL-15 cDNA, are vaccinated on day 2 with chicken vaccines and

11

then challenged on day 7 with the corresponding virulent viruses. They are observed for 8 weeks for signs of disease. It is expected that chicks injected with the pcDNA1 vector containing IL-15 cDNA will exhibit a reduced incidence of disease relative to controls.

Finally, IL-15 protein purified by the procedure described in Example 3 is administered intramuscularly to chicks at hatching, followed by a single daily administration on each of the following four days. Chicks are divided into three groups, receiving 0.01, 0.1 or 1.0 μ g per injection per day. A fourth group receives placebo injections. At hatching all chicks are vaccinated with chicken vaccines and then challenged on day 7 with the corresponding virulent viruses. They are then observed for 8 weeks for signs of disease. It is expected that chicks injected with IL-15 will exhibit a reduced incidence of disease relative to controls.

12

EXAMPLE 5

Use of Avian IL-15 in Growth Promotion

5 Mammalian IL-15 stimulates muscle growth (Quinn, L. S., *Endocrin.*, 136:3669, 1995) and semi-pure chicken IL-2 stimulates chicken body weight and increases feed conversion (U.S. Pat. No. 5,028,421). To evaluate the growth-promoting activity of avian IL-15, the methods described in Example 4 above may be used to administer IL-15 cDNA in a viral or plasmid vectors recombinant IL15 protein. Experimental and control chicks are monitored for weight gain and feed conversion for a period of six weeks. It is expected that one or more of these protocols will enhance chicken growth over controls.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 2

<210> SEQ ID NO 1

<211> LENGTH: 747

<212> TYPE: DNA

<213> ORGANISM: Gallus domesticus

<400> SEQUENCE: 1

```
cagataactg ggacactgcc atgatgtgca aagtactgat ctttggtgtg atttcggtag    60
caacgctaag gactacagct tatggagcat ctctatcatc agcaaaaagg aaacctcttc    120
aaacattaat aaaggattta gaaatattgg aaaatatcaa gaacaagatt catctcgagc    180
tctacacacc aactgagacc caggagtgcg ccagcaaac tctgcagtgt tacctgggag    240
aagtgggtac tctgaagaaa gaaactgaag atgacactga aattaaagaa gaatttgtaa    300
ctgctattca aaatatcgaa aagaacctca agagtcttac gggctctaag cacaccggaa    360
gtgaatgcaa gatctgtgaa gtaacaaca agaaaaaatt tcctgatttt ctccatgaac    420
tgaccaactt tgtgagatat ctgcaaaaat aagcaactaa tcatttttat tttactgcta    480
tgttatttat ttaattatit aattacagat aatttatata tttatcccg tggctaacta    540
atctgtgtgc cattctggga ccactgtatg ctcttagtct gggtgatatg acgtctgttc    600
taagatcata tttgatcctt tctgtaacct acgggctcaa aatgtacgtt ggaaaactga    660
ttgattctca ctttgcgtgt aaagtgatat gtgtttactg aaagaatttt taaaagtcac    720
ttctagatga catttaataa atttcag                                     747
```

<210> SEQ ID NO 2

<211> LENGTH: 143

<212> TYPE: PRT

<213> ORGANISM: Gallus domesticus

<400> SEQUENCE: 2

```
Met Met Cys Lys Val Leu Ile Phe Gly Cys Ile Ser Val Ala Thr Leu
1           5           10          15
Met Thr Thr Ala Tyr Gly Ala Ser Leu Ser Ser Ala Lys Arg Lys Pro
20          25          30
Leu Gln Thr Leu Ile Lys Asp Leu Glu Ile Leu Glu Asn Ile Lys Asn
35          40          45
Lys Ile His Leu Glu Leu Tyr Thr Pro Thr Glu Thr Gln Glu Cys Thr
50          55          60
```

-continued

Gln	Gln	Thr	Leu	Gln	Cys	Tyr	Leu	Gly	Glu	Val	Val	Thr	Leu	Lys	Lys
65					70				75					80	
Glu	Thr	Glu	Asp	Asp	Thr	Glu	Ile	Lys	Glu	Glu	Phe	Val	Thr	Ala	Ile
			85					90						95	
Gln	Asn	Ile	Glu	Lys	Asn	Leu	Lys	Ser	Leu	Thr	Gly	Leu	Asn	His	Thr
			100				105						110		
Gly	Ser	Glu	Cys	Lys	Ile	Cys	Glu	Ala	Asn	Asn	Lys	Lys	Lys	Phe	Pro
		115				120					125				
Asp	Phe	Leu	His	Glu	Leu	Thr	Asn	Phe	Val	Arg	Tyr	Leu	Gln	Lys	
	130					135				140					

What is claimed is:

1. An isolated nucleic acid which:
 - (a) comprises a nucleic acid sequence having at least 70% sequence homology, determined by a BLAST algorithm, to the sequence set forth in nucleotides 87-449 of SEQ ID NO:1; and
 - (b) encodes a polypeptide capable of stimulating thymidine incorporation in mitogen activated avian T-cells.
2. The complement of a nucleic acid according to claim 1.
3. An isolated nucleic acid according to claim 1 or 2, which nucleic acid is an avian nucleic acid isolated from chicken.
4. A vector construct comprising the nucleic acid of claim 1.
5. The vector construct according to claim 4, in which said nucleic acid is operatively associated with a promoter element capable of expressing the nucleic acid in a host cell.
6. The vector construct according to claim 4, in which the construct is a recombinant virus.
7. The vector construct according to claim 6, in which the recombinant virus is a turkey herpes virus or a fowl pox virus.
8. An isolated nucleic acid which:
 - (a) hybridizes to the full length of a nucleic acid having the complementary sequence of nucleotides 87-449 in SEQ ID NO:1 under conditions comprising (i) hybridization in 6xSSC and 0.5% SDS, and (ii) washing at 68° C. in 0.1xSSC and 0.5% SDS; and
 - (b) encodes a polypeptide capable of stimulating thymidine incorporation in mitogen activated avian T-cells.
9. The complement of a nucleic acid according to claim 8.
10. An isolated nucleic acid according to claim 8 or 9, which nucleic acid is an avian nucleic acid isolated from chicken.
11. An isolated nucleic acid which:
 - (a) hybridizes to the full length of a nucleic acid having the complementary sequence of nucleotides 87-449 in SEQ ID NO:1 under conditions comprising (i) hybridization in 6xSSC and 0.5% SDS, and (ii) washing at room temperature in 2xSSC and 0.5% SDS; and
 - (b) encodes a polypeptide capable of stimulating thymidine incorporation in mitogen activated avian T-cells.
12. The complement of a nucleic acid according to claim 11.
13. An isolated nucleic acid according to claim 11 or 12, which nucleic acid is an avian nucleic acid isolated from chicken.
14. A vector construct comprising the nucleic acid of claim 8 or 11.
15. A vector construct according to claim 14, in which said nucleic acid is operatively associated with a promoter element capable of expressing the nucleic acid in a host cell.
16. The vector construct according to claim 14, in which the construct is a recombinant virus.
17. The vector construct according to claim 16, in which the recombinant virus is a turkey herpes virus or a fowl pox virus.
18. An isolated nucleic acid having an open reading frame that encodes a polypeptide comprising the sequence of amino acid residues 23-143 set forth in SEQ ID NO:2 (FIG. 2).
19. The complement of a nucleic acid according to claim 18.
20. An isolated nucleic acid according to claim 18 or 19, which nucleic acid is an avian nucleic acid isolated from chicken.
21. An isolated nucleic acid according to claim 18, wherein the polypeptide comprises the amino acid sequence set forth in SEQ ID NO:2 (FIG. 2).
22. The complement of a nucleic acid according to claim 21.
23. An isolated nucleic acid according to claim 21 or 22 which nucleic acid is an avian nucleic acid isolated from chicken.
24. A vector construct comprising the nucleic acid of claim 18 or 21.
25. The vector construct according to claim 24, in which said nucleic acid is operatively associated with a promoter element capable of expressing the nucleic acid in a host cell.
26. The vector construct according to claim 24, in which the construct is a recombinant virus.
27. The vector construct according to claim 26, in which the recombinant virus is a turkey herpes virus or a fowl pox virus.

* * * * *



US006395306B1

(12) **United States Patent**
Cui et al.(10) **Patent No.:** **US 6,395,306 B1**
(45) **Date of Patent:** **May 28, 2002**(54) **BEE VENOM PROTEIN AND GENE
ENCODING SAME**(75) **Inventors:** **Xiangmin Cui, Cupertino; Yuefeng Lu,**
San Carlos, both of CA (US)(73) **Assignee:** **Pan Pacific Pharmaceuticals, Inc.,**
Lincoln, RI (US)(*) **Notice:** Subject to any disclaimer, the term of this
patent is extended or adjusted under 35
U.S.C. 154(b) by 0 days.(21) **Appl. No.:** **09/394,630**(22) **Filed:** **Sep. 13, 1999****Related U.S. Application Data**(60) **Provisional application No. 60/100,172, filed on Sep. 14,**
1998.(51) **Int. Cl.⁷** **A61K 35/64; A61K 35/24;**
C12P 21/06; C12P 2/02; C12N 15/00(52) **U.S. Cl.** **424/539; 424/537; 435/69.1;**
435/69.5; 435/320.1; 435/325; 536/23.1(58) **Field of Search** **435/69.1, 69.5,**
435/320.1, 325; 424/537, 539; 536/23.1(56) **References Cited****U.S. PATENT DOCUMENTS**

3,856,936 A 12/1974 Vick et al.

FOREIGN PATENT DOCUMENTS

WO WO 98/56400 12/1998

OTHER PUBLICATIONSBanks, et al., "Chemistry and Pharmacology of Honey-bee
Venom" in *Venoms of the Hymenoptera: Biochemical, Phar-*
macological and Behavioral Aspects (Piek, T., Ed.), Aca-
demic Press, chap. 7, pp. 397-398 (1986).Shipolini, Rudolf A., "Biochemistry of Bee Venom" in
Handbook of Natural Toxins, Allergans and Other Inverte-
brate Venoms, (Tu, Anthony T., Ed.) Marcel Dekker, Inc.,
vol. 2, pp. 58-59 (1984).Kim, Christopher M., M.D., Honey Bee Venom Therapy for
Arthritis (RA, OA) Fibromyositis (FM) and Peripheral Neu-
ritis (PN), *The Journal of the Korean Pain Research Society*,
Dec., 1991, vol. 1, No. 1, pp. 55-65.Product Information on Nectar Ease, [http://www.hugger-
s.com/NectarEase/productinfo.htm](http://www.hugger-
s.com/NectarEase/productinfo.htm), Jun. 15, 1998, pp. 1-8.
Freeman, Karen, Charles Mraz, 94, Advocate of Therapeutic
Bee Sting, Dies, *The New York Times, Sunday National*
Desk, Sep. 19, 1999, (as found at <http://nytimes.qpass.com>
pp. 1-4).Shkenderov, Stefan and Koburova, Krasimira; Adolapin—A
Newly Isolated Analgetic And Anti-Inflammatoy Polypep-
tide From Bee Venom, *Toxicon*, Great Britain, 1982, vol. 20,
No. 1, pp. 317-321.Koburova, K. L.; Michailova, S. G. and Shkenderov, S. V.;
Further Investigation on the Antiinflammatory Properties of
Adolapin—Bee Venom Polypeptide, *Bulgarian Academy of*
Sciences, 1985, vol. 8, No. 2, pp. 50-55.Koburova, K.; Mihailova, S. and Shkenderov, S.; The Anti-
pyretic Effect Of One Polypeptide From The Bee Poison—
Adolapin, *Eksp Med Morfol*, 1984, 23(3) pp. 143-148.Cohen, A., M.D.; Dubbs, A. W., M.D.; Pearah, J. B., M.D.;
and Best, C. J., M.D.; Bee Venom in the Treatment of
Chronic Arthritis, A Comparative Study, *The Pennsylvania*
Medical Journal, Jun., 1942, vol. 45, No. 9, pp. 957-959.Marz, Charles, Bee Venom for Arthritis—An Update,
American Bee Journal, Feb., 1982, pp. 121-123.Shkenderov, S., New Pharmacobiochemical Data on the
Anti-Inflammatory Effect of Bee Venom, in *Animal, Plant,*
and Microbial Toxins, 1976, vol. 2, pp. 319-336.Hollander, J. L., M.D., Bee Venom In The Treatment Of
Chronic Arthritis, *The American Journal OF The Medical*
Sciences, Jun., 1941, vol. 201, No. 6, pp. 796-801.Guyton, F. E., Bee Sting Therapy for Arthritis and Neuritis,
Journal of Economic Entomology, 1947, vol. 40, pp.
469-472.Kroner, J., M.D.; Lintz, R.M., M.D.; Tyndall, M., M.D.,
Andersen, L., M. D.; and Nicholls, E. E., M.D.; The Treat-
ment of Rheumatoid arthritis With An Injectable Form Of
Bee Venom, *Annals Of Internal Medicine*, Jan., 1938, vol. 11,
No. 7, pp. 1077-1083.Chang, Yi-Han and Bliven, Marcia L.; Anti-Arthritic Effect
of Bee Venom, *Agents and Actions*, 1979, vol. 9/2, pp.
205-211.Billingham, M. E. J.; Morley, J.; Hanson, J. M.; Shipolini, R.
A.; and Vernon, C. A.; An Anti-Inflammatory Peptide from
Bee Venom, *Nature*, Sep. 21, 1973, vol. 245, pp. 163-164.
Zurier, R. B.; Mitnick, H.; Bloomgarden, D.; and Weiss-
mann, G.; Effect of Bee Venom on Experimental Arthritis,
Annal of the Rheumatic Diseases, 1973, vol. 32, pp.
466-470.Vick, J. A.; Warren, G. B.; and Brooks, R. B., Jr.; The Effect
Of Treatment With Whole Bee Venom On Cage Activity
And Plasma Cortisol In The Arthritic Dog, *Inflammation*,
1975-1976, vol. 1, No. 2, pp. 167-174.Somerfield, S. D.; Stach, J. L.; Mraz, D.; Gervais, F.; and
Skamene, E.; Bee Venom Inhibits Superoxide Production By
Human Neutrophils, *Inflammation*, 1984, vol. 8, No. 4, pp.
385-391.Shkenderov, S.; A Protease Inhibitor In Bee Venom (Iden-
tification, Partial Purification and some Properties), *FEBS*
Letters, Jul. 1973, vol. 33, No. 3, pp. 343-347.Hyre, H. M. and Smith, R. A.; Immunological Effects Of
Honey Bee (APIS Mellifera) Venom Using BALB/c Mice,
Toxicon, 1986, vol. 24, No. 5, pp. 435-440.

(List continued on next page.)

Primary Examiner—Jeffrey Stucker**Assistant Examiner**—Jegatheesan Scharaseyon(74) **Attorney, Agent, or Firm**—Townsend and Townsend
and Crew LLP; Scott Ausenhus(57) **ABSTRACT**The invention provides a novel protein, PX3.101, which can
be isolated from honey bee venom, antibodies against the
polypeptide and nucleic acids encoding PX3.101 and frag-
ments thereof. The invention also provides pharmaceutical
compositions based upon PX3.101 polypeptide and methods
for using same in the treatment of various diseases, includ-
ing various inflammatory diseases such as rheumatoid arthri-
tis.**24 Claims, 9 Drawing Sheets**

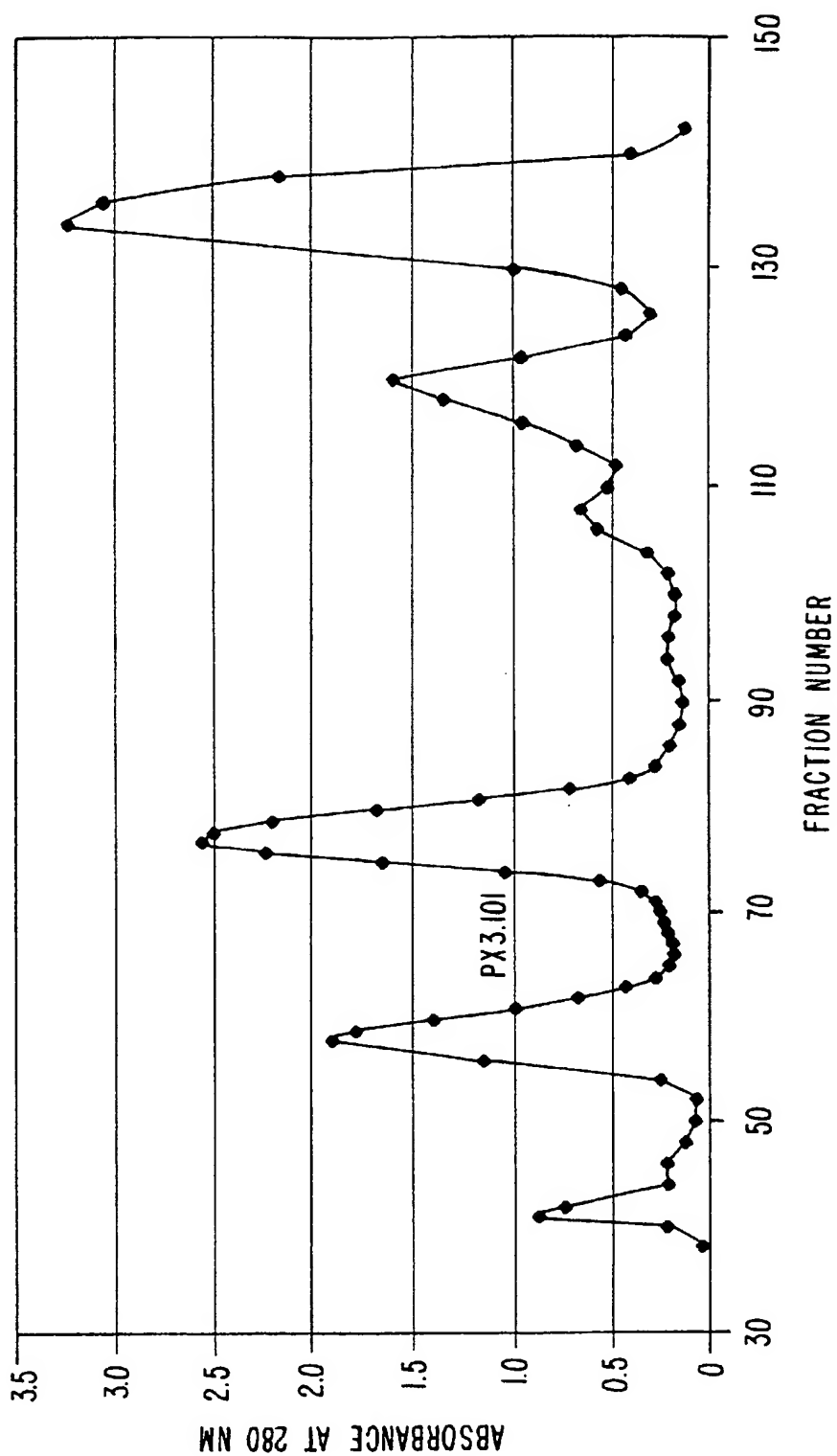


FIG. 1.

FIG. 2A.

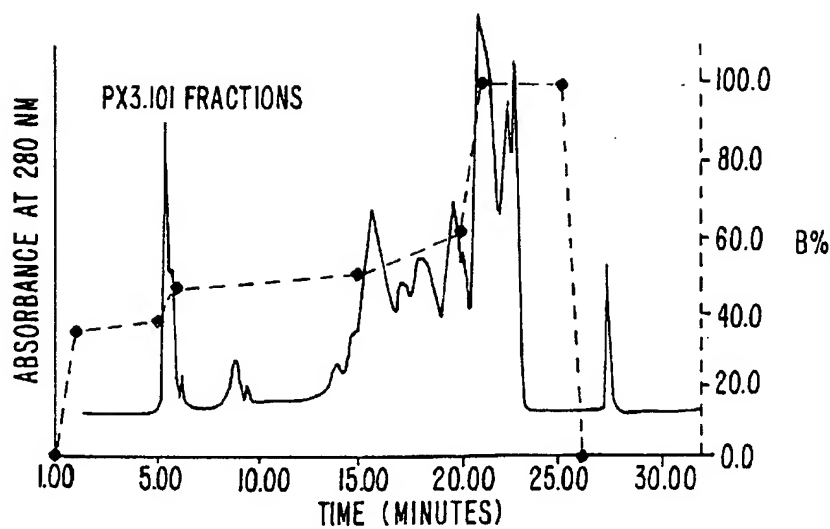


FIG. 2B.

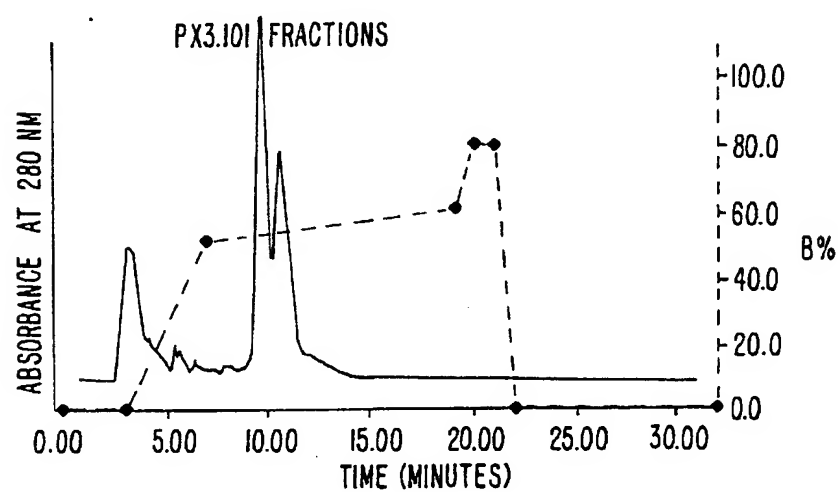
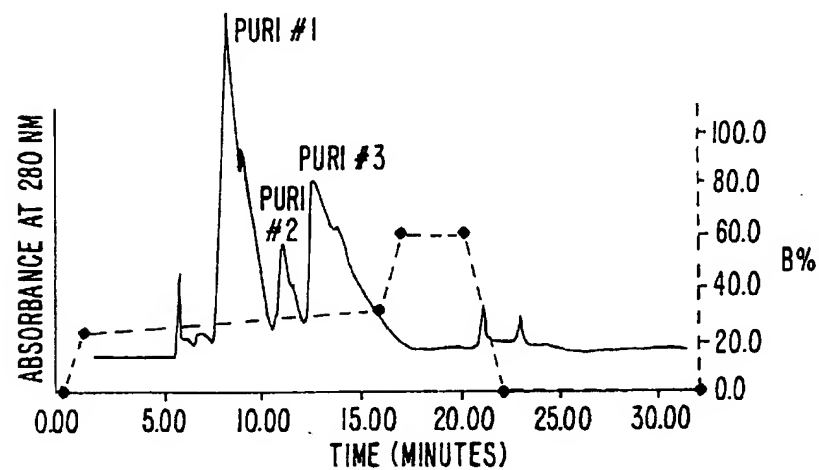
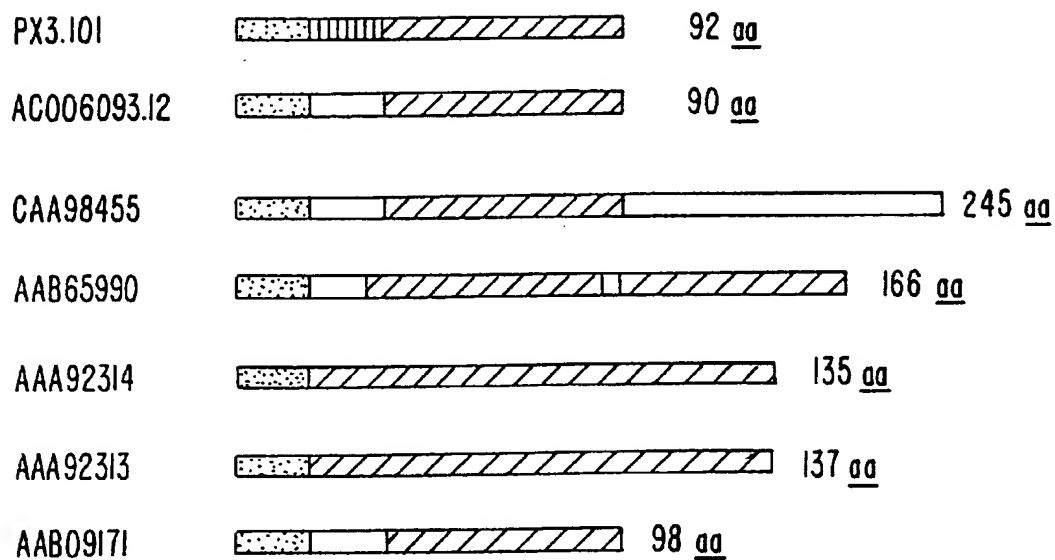


FIG. 2C.





SIGNAL PEPTIDE



GGX REPEATS



CYSTEINE-RICH REGION

aa

AMINO ACID

FIG. 3C.

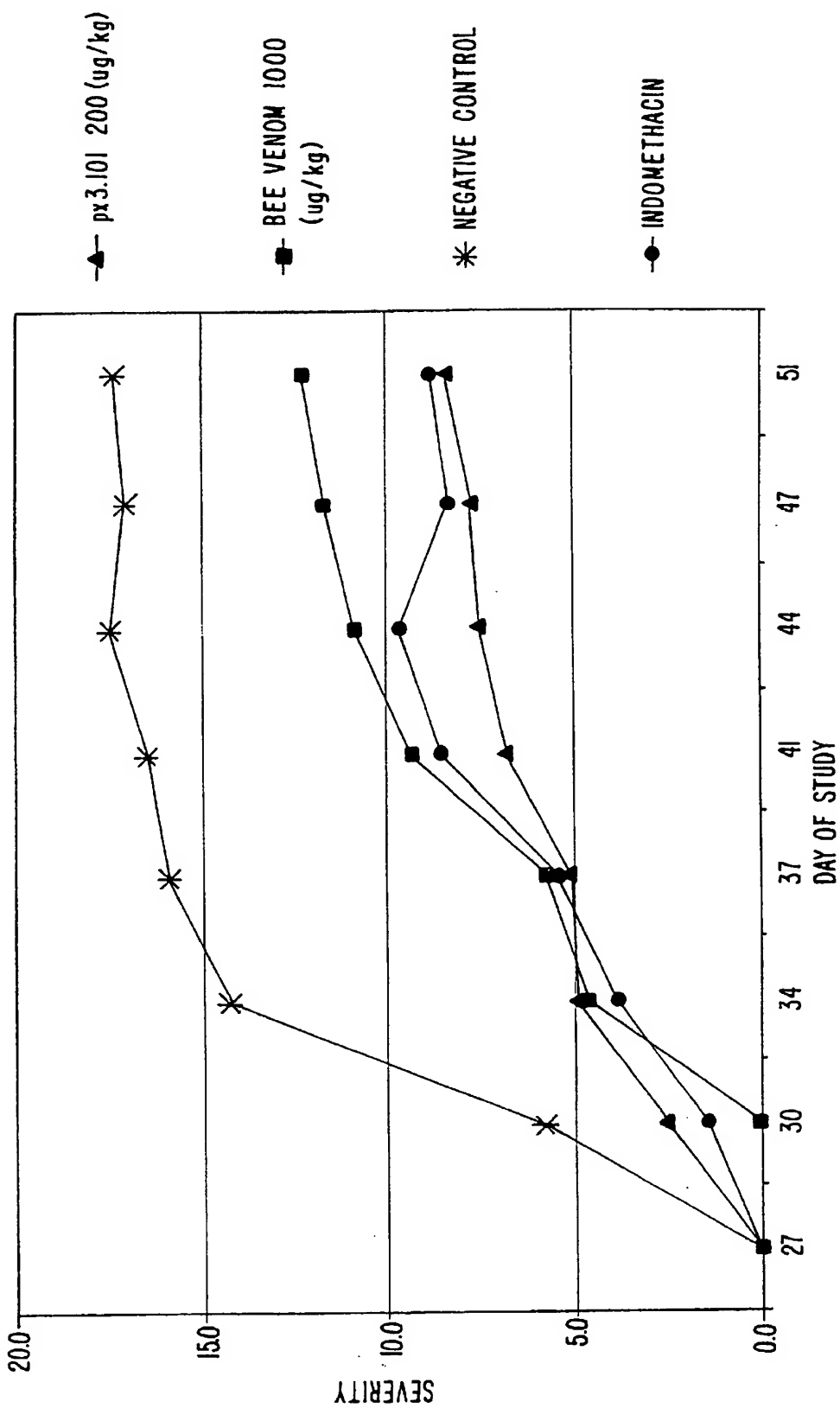


FIG. 4.



FIG. 5A

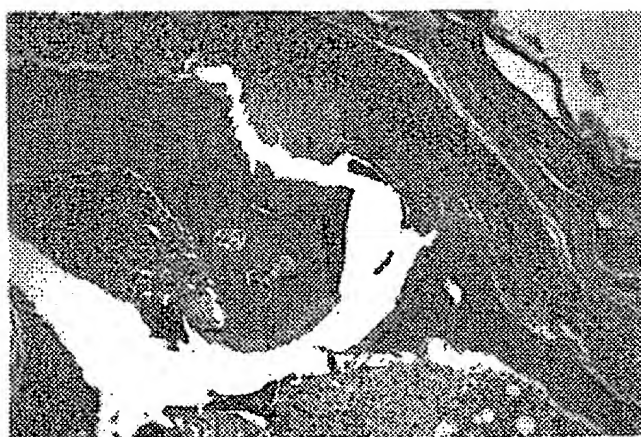


FIG. 5B

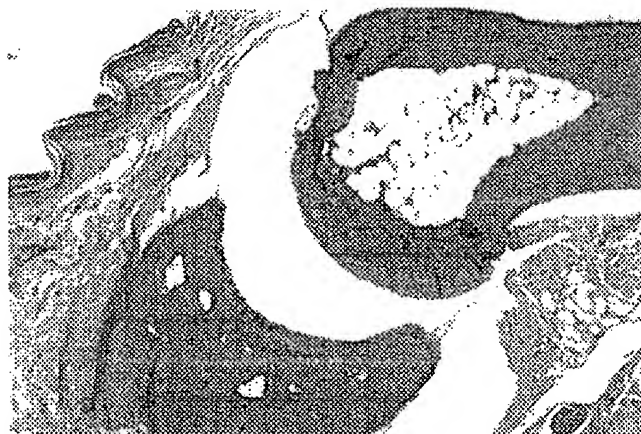


FIG. 5C

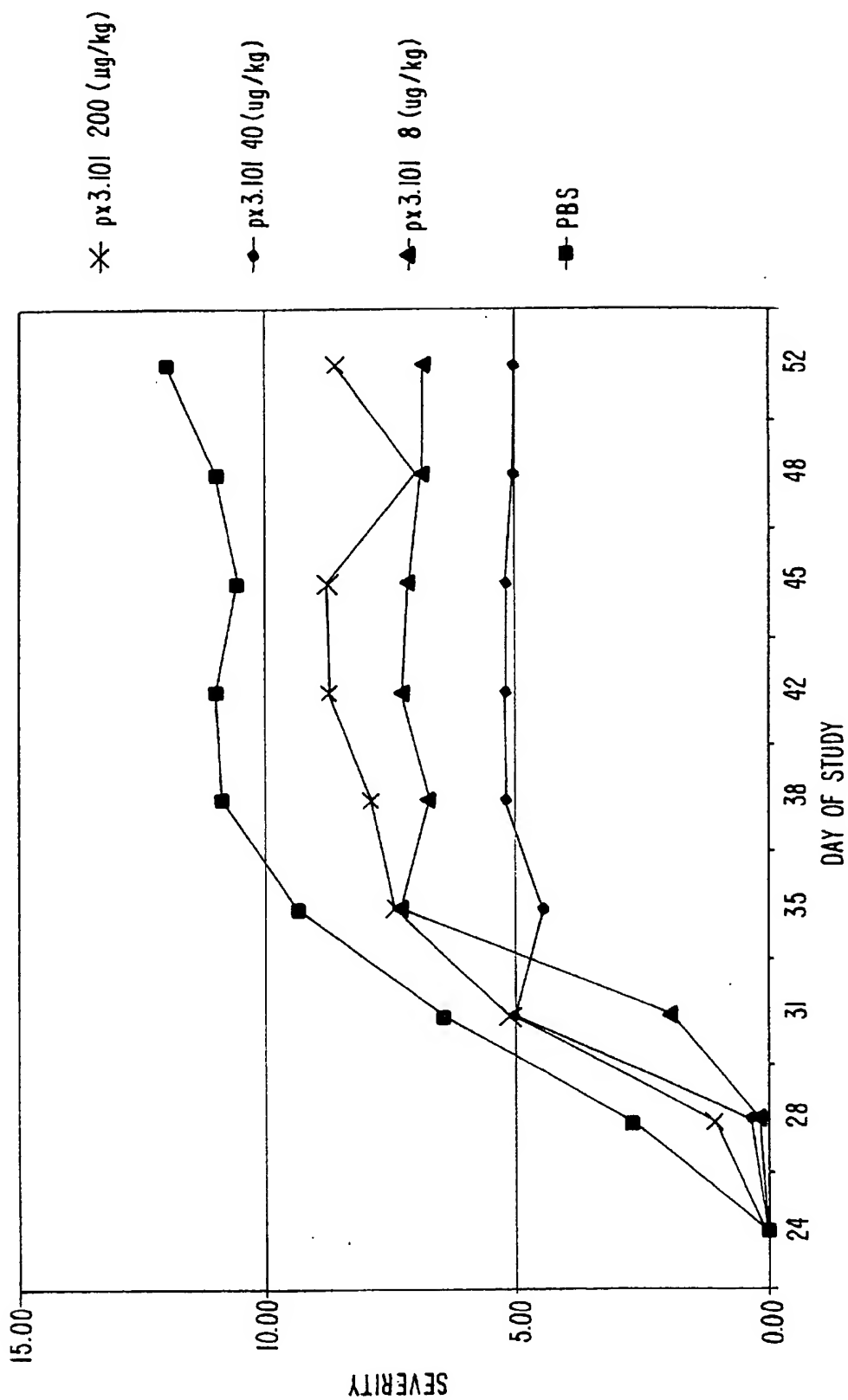
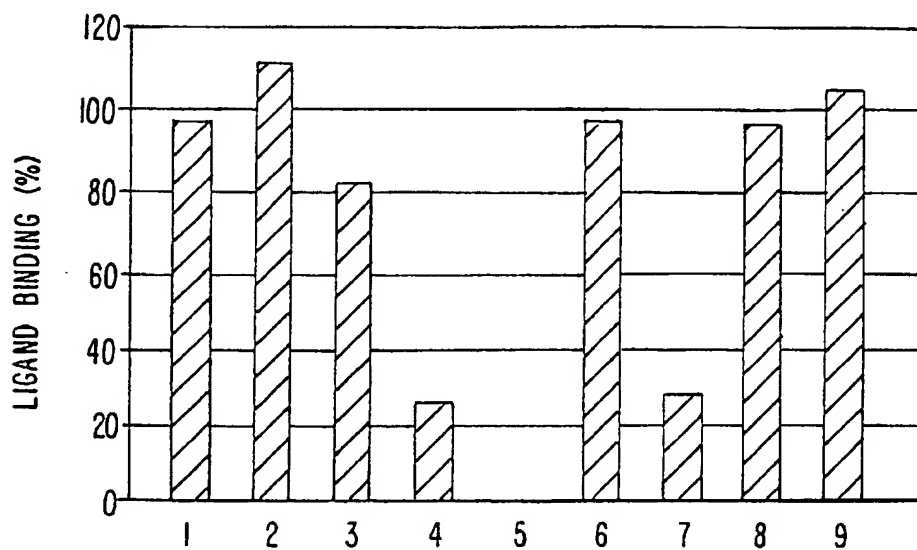
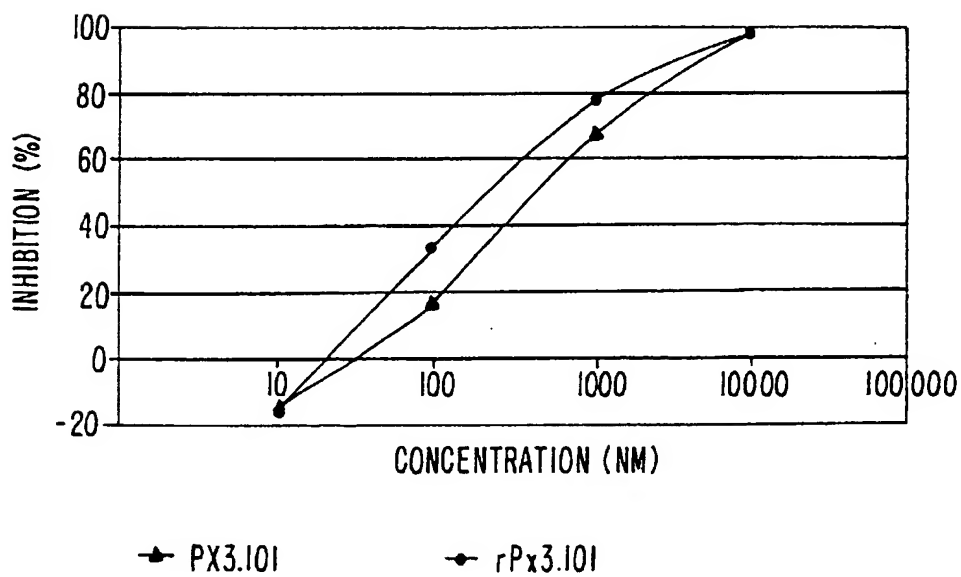
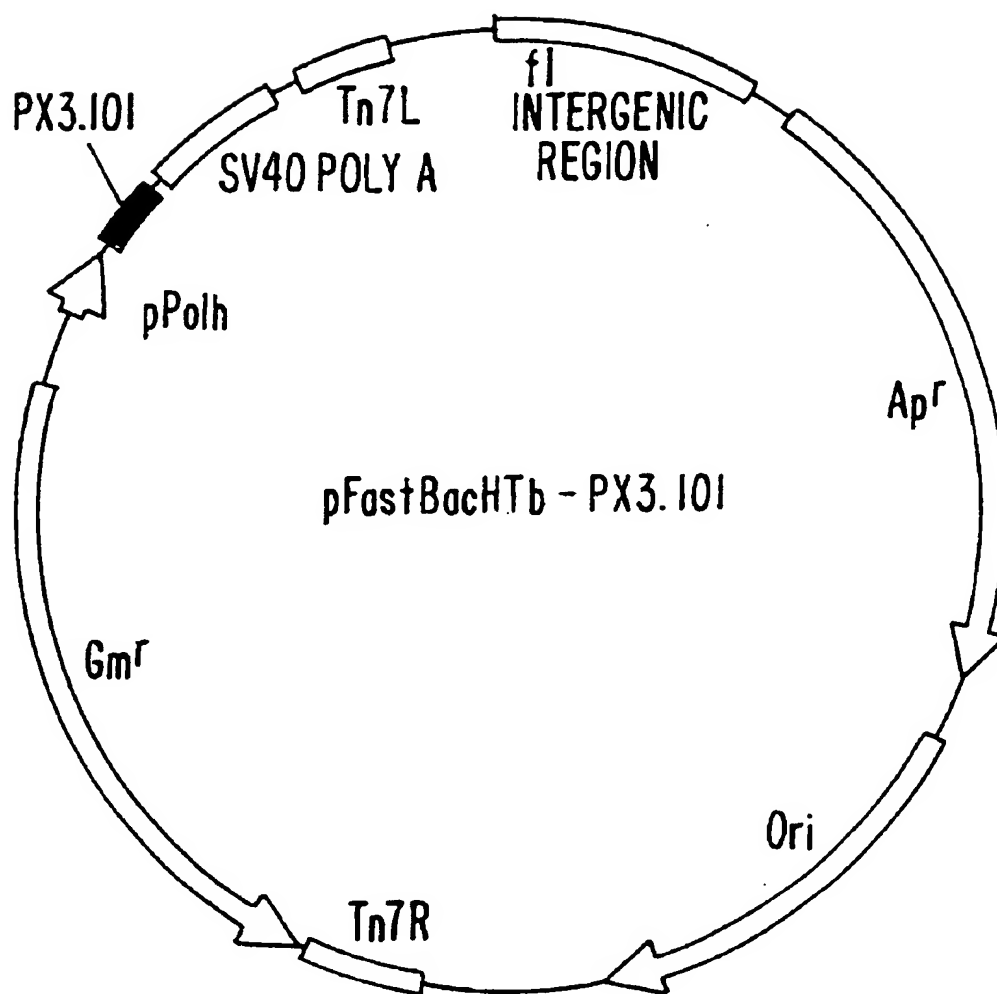


FIG. 6.

*FIG. 7A.**FIG. 7B.*

**FIG. 8.**

1

BEE VENOM PROTEIN AND GENE ENCODING SAME

CROSS REFERENCES TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 60/100,172, filed Sep. 14, 1998, which is incorporated herein by reference in its entirety for all purposes.

FIELD OF THE INVENTION

This invention relates to the field of cloning and expression of a protein with therapeutic value in the treatment of various diseases, especially inflammatory diseases such as rheumatoid arthritis. More specifically, the invention relates to a novel protein called PX3.101 purified from honey bee venom and the gene encoding the protein.

BACKGROUND OF THE INVENTION

The immune system plays a critical beneficial role in combating infections. However, in some instances improper immune responses can result in many disabling diseases. Autoimmune or immune-system mediated diseases may be either B-cell mediated (i.e., antibody-mediated) or T-cell mediated. Many autoimmune diseases involve an undesirable inflammatory response. Examples of such diseases include rheumatoid arthritis, chronic hepatitis, Crohn's disease, psoriasis, vasculitis, and the like.

Existing therapies for autoimmune diseases, particularly those involving an undesirable inflammatory response are inadequate. Most immune system-mediated diseases are chronic conditions that require the prolonged administration of drugs to address the symptoms of the disease. Accordingly, an important criterion for drugs used to treat these diseases is low toxicity. However, many drugs utilized to treat autoimmune diseases (e.g., steroids and non-steroidal anti-inflammatory compounds (NSAIDs)), have significant toxic side effects that become manifest after prolonged periods of use. Various immunosuppressive drugs (e.g., cyclosporin A and azathioprine) have also been used to treat autoimmune diseases. However, these compounds are relatively non-specific and have the adverse effect of weakening the entire immune system, thus leaving the patient susceptible to infectious disease.

A variety of inflammatory diseases, including rheumatoid arthritis, are associated with interleukin 8 (IL-8). IL-8 is a chemokine that promotes the recruitment and activation of neutrophil leukocytes and represents one of several endogenous mediators of acute inflammatory response. IL-8 has also been variously referred to as neutrophil-activating factor, monocyte-derived neutrophil chemotactic factor, interleukin-8 (IL-8) and neutrophil-activating peptide. The term IL-8 has gained the most widespread acceptance and is used herein.

Inflammation and autoimmune responses commence with the migration of leukocytes out of the microvascular into the extravascular space in response to chemoattractant molecules. Chemoattractants may originate from the host and include chemokines and activated complement components, or may be released from an invading organism. Once exposed to chemoattractants within the vasculature, the leukocytes become activated and capable of adhering to the endothelium providing the first step in the development of inflammation. Stimulated neutrophils adhere to the endothelium of the microvasculature in response to a gradient of

2

chemoattractants which direct the cells into the extravascular space toward the source of the chemoattractant. See, for example, Anderson et al., *Journal Clin. Invest.* 74:536-551, (1984); Ley, K. et al., *Blood* 77:2553-2555, (1991); Paulson, J. C., "Selectin-carbohydrate-mediated adhesion of leukocytes", *Adhesion: Its Role in Inflammatory Disease*, W. H. Freeman, 1992; Lasky, L. A., "The homing receptor" (LECAM 1/L-selectin), *Adhesion: Its role in inflammatory disease*, W. H. Freeman, (1992).

Rheumatoid arthritis is one of the more prevalent autoimmune and inflammatory diseases. The disease afflicts approximately 1% of the total population and about 2.5 million persons in the United States alone. Direct prescription usage is estimated at \$5.6 billion worldwide. For individuals suffering from rheumatoid arthritis, the individual's immune system mistakenly perceives the body's own joint tissue as foreign and thus initiates an abnormal immune response. The disease is characterized by chronic inflammation, destruction of cartilage, and ultimately bone erosion and the destruction of joints.

As with other inflammatory diseases, known treatments for IL-8 mediated diseases and rheumatoid arthritis can include the use of nonspecific immunosuppressive drugs that suppress the entire immune system; as noted above, however, such treatments put the patient at risk for contracting an infectious disease. Prolonged use of such drugs can also result in severe side effects. Moreover, immunosuppressive drugs are only partially effective in mitigating the symptoms of rheumatoid arthritis and the utility of the treatment tends to decrease with time.

Other therapies currently used are non-steroid anti-inflammatory drugs (NSAIDs), corticosteroids and a variety of disease modifying anti-rheumatic drugs (DMARDs). There is general dissatisfaction with these drugs for two major reasons: (i) incidence of adverse side effects, which lead to over 700, 000 hospitalizations every year, and (ii) inability to reverse disease progression.

Given the paucity of effective treatments for inflammatory diseases and autoimmune diseases generally, and the need for effective compositions for treating diseases associated with IL-8 such as rheumatoid arthritis more particularly, there is a significant need for new substances that can be used in the treatment of these diseases.

The present invention provides novel isolated proteins and nucleic acids encoding the proteins that are effective in treating autoimmune and inflammatory diseases, especially rheumatoid arthritis. The peptides of the invention also can inhibit the binding of IL-8 to its receptor and inhibit a variety of enzymes associated with inflammatory diseases.

SUMMARY OF THE INVENTION

The invention provides nucleic acid molecules that include a polynucleotide sequence that encodes a PX3.101 polypeptide or fragments thereof. The polypeptides of the invention have an amino acid sequence at least 75% identical to an amino acid sequence as set forth in SEQ ID NO:2 over a region at least about 40 amino acids in length when compared using the BLASTP algorithm with a wordlength (W) of 3, and the BLOSUM62 scoring matrix. The polynucleotide sequences are preferably at least 75% identical to a nucleic acid sequence set forth in residues 74 to 349 of SEQ ID NO:1 over a region of at least 50 nucleotides in length when compared using the BLASTN algorithm with a wordlength (W) of 11, M=5, and N=-4.

Nucleic acids of the invention also include isolated nucleic acid molecules comprising a nucleotide sequence

selected from the group consisting of: (a) deoxyribonucleotide sequence complementary to nucleotides 74 to 349 of SEQ ID NO:1; (b) a ribonucleotide sequence complementary to nucleotides 74 to 349 of SEQ ID NO:1; (c) a nucleotide sequence complementary to the deoxyribonucleotide sequence of (a) or to the ribonucleotide sequence of (b); (d) a nucleotide sequence of at least 23 consecutive nucleotides capable of hybridizing to nucleotides 74 to 349 of SEQ ID NO:1; and (e) a nucleotide sequence capable of hybridizing to a nucleotide sequence of (d). The nucleic acid molecules of the invention will generally hybridize to a polynucleotide sequence consisting of nucleotides 74 to 349 of SEQ ID NO:1 under stringent conditions. An exemplary nucleic acid of the invention is a nucleic acid consisting of nucleotides 74 to 349 of SEQ ID NO:1. Nucleic acids of the invention also include those which are capable of being amplified with forward primer 5' AAGGATCCACAGTCAACGTAAGTTC 3' (SEQ ID NO:3) and reverse primer 5' ACTGATAAAATAATAAC 3' (SEQ ID NO:5).

The invention also provides polypeptides that have an amino acid sequence at least 75% identical to an amino acid sequence as set forth in SEQ ID NO:2 over a region at least 40 amino acids in length when compared using the BLASTP algorithm with a wordlength (W) of 3, and the BLOSUM62 scoring matrix. Polypeptides of the invention include polypeptides encoded by a nucleic acid segment that hybridizes under stringent conditions to a nucleic acid fragment having the sequence set forth in SEQ ID NO:1. Polypeptides that are also included are those having an antigenic determinant common to a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2. An example of a polypeptide of the invention is a polypeptide having the sequence set forth in SEQ ID NO:2. The invention further provides polypeptide fragments that include at least 12 contiguous amino acids from SEQ ID NO:2. Other polypeptides provided by the invention are purified polypeptides which include a signal peptide, at least 3 GGX repeats, and a C terminal segment extending from the last GGX repeat to the C-terminus which contains at least 7 cysteine residues, wherein X is any amino acid and the polypeptide is less than 140 amino acids in length.

The invention also includes cells that include a vector containing a nucleic acid of the invention. For example, the invention provides cells that have a recombinant expression cassette containing a promoter operably linked to a polynucleotide sequence which encodes a polypeptide as described herein. Both prokaryotic and eukaryotic cells that express polypeptides of the invention are provided.

Methods for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:2 or fragments thereof are also provided. The methods generally include culturing a host cell containing a recombinant expression cassette under conditions suitable for the expression of the polypeptide and then recovering the polypeptide from the host cell culture.

The invention further provides antibodies that are specific for the polypeptides and polypeptide fragments of the invention.

A variety of pharmaceutical compositions are provided by the invention. These compositions typically contain a polypeptide as described herein and a pharmaceutically acceptable excipient. In some instances, the compositions also include a complementary agent which is known to be effective in treating inflammatory diseases. Various compositions can be used to treat various diseases, including, for example, inflammatory diseases, cancer, autoimmune

diseases, pain, and diseases associated with chemokine imbalances. These methods generally involve administering a therapeutically effective dose of one of the pharmaceutical compositions of the invention to a patient suffering from a disease.

The invention further provide methods for inhibiting the interaction between certain chemokines with their receptors or for inhibiting enzymes associated with inflammatory diseases. These method generally involve admixing a polypeptide of the invention with a solution containing a chemokine and its receptor or a solution containing an enzyme associated with inflammatory diseases.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 shows an elution profile for a honey bee venom suspension eluted through a Sephadex G-50 sizing column. One ml of honeybee venom suspension (approximate 0.5 g solid material) was diluted in 10 ml of column buffer (ammonium formate buffer, 0.1 M, pH 4.6), spun down and filtered through a 0.45 μ m filter. The resulting solution was loaded onto a Sephadex G-50 column (two connected columns, each 1.5x170 cm (diameterxheight)) pre-equilibrated with the column buffer. The column was eluted at about 0.6 ml/min, and fractions of 100 drops (approximately 4.0 ml) were collected.

FIG. 2 shows elution profiles for the three HPLC purification steps used to purify PX3.101 protein.

FIG. 2A is an HPLC elution profile for fractions obtained from the G-50 sizing column. Fractions from the G-50 sizing columns such as shown in FIG. 1 were tested for the presence of PX3.101 by SDS-PAGE. Positive fractions were pooled and loaded onto a Reverse Phase (RP) HPLC column (semi-prep C-18 column). The column was eluted with an acetonitrile gradient (see Reverse Phase HPLC section in Example 1 for detailed information). PX3.101-containing fractions were collected and freeze-dried.

FIG. 2B is an elution profile showing further purification of PX3.101 by ion exchange HPLC in which PX3.101 powder from the first RP-HPLC chromatography purification step was dissolved in 0.1 M ammonium formate (pH 5.8) and loaded onto an ion exchange HPLC column (see Example 1 for details).

FIG. 2C is an example of the elution profile for the final purification step of PX3.101 using a second RP-HPLC column. PX3.101 fractions from the ion exchange HPLC purification shown in FIG. 2B were pooled and loaded onto another RP-HPLC column. Shown is the chromatography of the purification of PX3.101 from 2 g dry honeybee venom. PX3.101 fractions (Puri-#1, Puri-#2 and Puri-#3) are indicated on the profile. The mixture of Puri-1, Puri-2 and Puri-3 was used for animal studies and mechanism studies. The differences between PX3.101 fractions Puri-#1, Puri-#2 and Puri-93 are discussed in the text.

FIG. 3A shows the full-length nucleotide sequence for the cDNA encoding full-length PX3.101 (SEQ ID NO:1) and the predicted protein sequence of PX3.101 (SEQ ID NO:2). The nucleotide sequence is the number listed on the left in plain type; the sequence begins with the first nucleotide of the PX3.101 cDNA. Amino acids are numbered in italics on the right. The in-frame stop codon is denoted by an asterisk. Four peptide sequences obtained from peptide sequencing are underlined.

FIG. 3B is a schematic representation of the PX3.101 protein structure showing the signal peptide region, the region containing Gly Gly Xaa (where Xaa=any amino acid and Gly=Glycine) repeats and the cysteine rich region.

FIG. 3C is a schematic representation of the structures of PX3.101 protein and its potential homologues. The number of the amino acids and accession numbers are included.

FIG. 4 is a chart showing the effectiveness of PX3.101 in controlling inflammation in the CIA (collagen-induced arthritis) mouse animal model. Indomethacin and bee venom serve as positive controls. PBS serves as the negative control. Severity is a measure of the degree of inflammation measured in each treatment group (see Example V for details).

FIG. 5 includes photographs of representative joint tissues from mice in different treatment groups.

FIG. 5A (normal control) shows a joint for a mouse that was not injected with collagen to induce rheumatoid arthritis but which was injected with phosphate buffer in normal saline (PBS).

FIG. 5B (negative control) is a photograph of the joint of a mouse which was injected with collagen to induce rheumatoid arthritis and also injected with PBS.

FIG. 5C is a photograph of the joint of a mouse which was injected with collagen to induce rheumatoid arthritis and also injected with a solution containing PX3.101 (200 $\mu\text{g/kg}$). Additional details are provided in Example V.

FIG. 6 is a chart showing the effectiveness of various concentrations of PX3.101 in controlling inflammation in the CIA (collagen-induced arthritis) mice model. Three doses of PX3.101 (8 $\mu\text{g/kg}$, 40 $\mu\text{g/kg}$, and 200 $\mu\text{g/kg}$) were tested. Severity is a measure of the degree of inflammation measured in each treatment group (see Example V for additional details).

FIG. 7A is a chart showing inhibition of binding between the chemokine IL-8 with its receptor CXCR2 at different concentrations of PX3.101 purified from honeybee venom. Purified PX3.101 was added to 0.2 ml reaction solution to give final PX3.101 concentrations of 0, 0.01, 0.1, 1.0 and 10 μM (Columns 1 through 5, respectively). Reaction mixtures also contained 0.15 mg/ml membrane preparation of human recombinant CHO cells expressing CXCR2, 0.015 nM ^{125}I IL-8, and 10 nM unlabeled IL-8, and were incubated for 60 minutes at room temperature. Bound radioligand was separated from unbound radioligand and the radioactivity counted on a gamma counter. The ligand bound was normalized and calculated as a percentage of the ligand bound in solutions without PX3.101. The effects of the purified PX3.101 on the binding of IL-8 to the receptor CXCR1 (Columns 6:0 μM ; and column 7:10 μM) and of TNF- α (tumor necrosis factor - α) to TNF- α receptor are also shown (Column 8:0 μM ; and column 9:10 μM). Details concerning the assay methods are described in Example VI. The results represent the averages of two measurements.

FIG. 7B shows inhibition plots demonstrating inhibition of IL-8/CXCR2 interaction by recombinant PX3.101 protein (●) and native PX3.101 protein from honeybee venom (▲).

FIG. 8 is a schematic of the expression vector constructed to produce recombinant PX3.101 protein.

DEFINITIONS

The term "nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally-occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence includes the complementary sequence thereof. A "subsequence" refers to a sequence of nucleotides or amino

acids that comprise a part of a longer sequence of nucleotides or amino acids (e.g., a polypeptide), respectively.

A "probe" is a nucleic acid capable of binding to a target nucleic acid of complementary sequence through one or more types of chemical bonds, usually through complementary base pairing, usually through hydrogen bond formation, thus forming a duplex structure. The probe binds or hybridizes to a "probe binding site." A probe may include natural (i.e. A, G, C, or T) or modified bases (7-deazaguanosine, inosine, etc.). A probe can be an oligonucleotide which is a single-stranded DNA. Oligonucleotide probes can be synthesized or produced from naturally occurring polynucleotides. In addition, the bases in a probe can be joined by a linkage other than a phosphodiester bond, so long as it does not interfere with hybridization. Thus, probes may be peptide nucleic acids in which the constituent bases are joined by peptide bonds rather than phosphodiester linkages (see, for example, Nielsen et al., *Science* 254, 1497-1500 (1991)). Some probes may have leading and/or trailing sequences of noncomplementarity flanking a region of complementarity.

The terms "polypeptide," "peptide" and "protein" are used interchangeably to refer to a polymer of amino acid residues. The term also applies to amino acid polymers in which one or more amino acids are chemical analogues of a corresponding naturally-occurring amino acids.

The term "operably linked" refers to functional linkage between a nucleic acid expression control sequence (such as a promoter, signal sequence, or array of transcription factor binding sites) and a second polynucleotide, wherein the expression control sequence affects transcription and/or translation of the second polynucleotide.

A "heterologous sequence" or a "heterologous nucleic acid," as used herein, is one that originates from a source foreign to the particular host cell, or, if from the same source, is modified from its original form. Thus, a heterologous gene in a prokaryotic host cell includes a gene that, although being endogenous to the particular host cell, has been modified. Modification of the heterologous sequence can occur, e.g., by treating the DNA with a restriction enzyme to generate a DNA fragment that is capable of being operably linked to the promoter. Techniques such as site-directed mutagenesis are also useful for modifying a heterologous nucleic acid.

The term "recombinant" when used with reference to a cell indicates that the cell replicates a heterologous nucleic acid, or expresses a peptide or protein encoded by a heterologous nucleic acid. Recombinant cells can contain genes that are not found within the native (non-recombinant) form of the cell. Recombinant cells can also contain genes found in the native form of the cell wherein the genes are modified and re-introduced into the cell by artificial means. The term also encompasses cells that contain a nucleic acid endogenous to the cell that has been modified without removing the nucleic acid from the cell; such modifications include those obtained by gene replacement, site-specific mutation, and related techniques.

A "recombinant expression cassette" or simply an "expression cassette" is a nucleic acid construct, generated recombinantly or synthetically, that has control elements that are capable of effecting expression of a structural gene that is operably linked to the control elements in hosts compatible with such sequences. Expression cassettes include at least promoters and optionally, transcription termination signals. Typically, the recombinant expression cassette includes at least a nucleic acid to be transcribed

(e.g., a nucleic acid encoding a desired polypeptide) and a promoter. Additional factors necessary or helpful in effecting expression can also be used as described herein. For example, an expression cassette can also include nucleotide sequences that encode a signal sequence that directs secretion of an expressed protein from the host cell. Transcription termination signals, enhancers, and other nucleic acid sequences that influence gene expression, can also be included in an expression cassette.

The term "isolated," "purified" or "substantially pure" means an object species (e.g., PX3.101 polypeptide or fragments thereof, or a nucleic acid fragment) is the predominant macromolecular species present (i.e., on a molar basis it is more abundant than any other individual species in the composition), and preferably the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present. Generally, an isolated, purified or substantially pure composition will comprise more than 80 to 90 percent of all macromolecular species present in a composition. Most preferably, the object species is purified to essential homogeneity (i.e., contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species.

The term "complementary" means that one nucleic acid is identical to, or hybridizes selectively to, another nucleic acid molecule. Selectivity of hybridization exists when hybridization occurs that is more selective than total lack of specificity. Typically, selective hybridization will occur when there is at least about 55% identity over a stretch of at least 14–25 nucleotides, preferably at least 65%, more preferably at least 75%, and most preferably at least 90%. Preferably, one nucleic acid hybridizes specifically to the other nucleic acid. See M. Kanehisa, *Nucleic Acids Res.* 12:203 (1984).

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptides, refer to two or more sequences or subsequences that are the same or have a specified percentage of nucleotides or amino acid residues that are the same, when compared and aligned for maximum correspondence, as measured using a sequence comparison algorithm such as those described below for example, or by visual inspection.

The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to two or more sequences or subsequences that have at least 75%, preferably at least 85%, more preferably at least 90%, 95% or higher nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using a sequence comparison algorithm such as those described below for example, or by visual inspection. Preferably, the substantial identity exists over a region of the sequences that is at least about 40–50 residues in length, preferably over a longer region than 50 amino acids, more preferably at least about 90–100 residues, and most preferably the sequences are substantially identical over the full length of the sequences being compared, such as the coding region of a nucleotide for example.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (see generally Ausubel et al., *supra*).

Another example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., *J. Mol. Biol.* 215:403–410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. For identifying whether a nucleic acid or polypeptide is within the scope of the invention, the default parameters of the BLAST programs are suitable. The BLASTN program (for nucleotide sequences) uses as defaults a word length (W) of 11, an expectation (E) of 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a word length (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix. The TBLASTN program (using protein sequence for nucleotide sequence) uses as defaults a word length (W) of 3, an expectation (E) of 10, and a BLOSUM 62 scoring matrix. (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873–5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to

each other under stringent conditions. "Bind(s) substantially" refers to complementary hybridization between a probe nucleic acid and a target nucleic acid and embraces minor mismatches that can be accommodated by reducing the stringency of the hybridization media to achieve the desired detection of the target polynucleotide sequence. The phrase "hybridizing specifically to", refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent conditions when that sequence is present in a complex mixture (e.g., total cellular DNA or RNA).

The term "stringent conditions" refers to conditions under which a probe will hybridize to its target subsequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5° C. lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. (As the target sequences are generally present in excess, at T_m, 50% of the probes are occupied at equilibrium). Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30° C. for short probes (e.g., 10 to 50 nucleotides) and at least about 60° C. for long probes (e.g., greater than 50 nucleotides). Stringent conditions can also be achieved with the addition of destabilizing agents such as formamide.

A further indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid, as described below. The phrases "specifically binds to a protein" or "specifically immunoreactive with," when referring to an antibody refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, a specified antibody binds preferentially to a particular protein and does not bind in a significant amount to other proteins present in the sample. Specific binding to a protein under such conditions requires an antibody that is selected for its specificity for a particular protein. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See, e.g., Harlow and Lane (1988) *Antibodies, A Laboratory Manual*, Cold Spring Harbor Publications, New York, for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity.

"Conservatively modified variations" of a particular polynucleotide sequence refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given polypeptide. For instance, the codons CGU, CGC, CGA, CGG, AGA, and AGG all encode the amino acid arginine. Thus, at every position where an arginine is

specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of "conservatively modified variations." Every polynucleotide sequence described herein which encodes a polypeptide also describes every possible silent variation, except where otherwise noted. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule by standard techniques. Accordingly, each "silent variation" of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

A polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. A "conservative substitution," when describing a protein, refers to a change in the amino acid composition of the protein that does not substantially alter the protein's activity. Thus, "conservatively modified variations" of a particular amino acid sequence refers to amino acid substitutions of those amino acids that are not critical for protein activity or substitution of amino acids with other amino acids having similar properties (e.g., acidic, basic, positively or negatively charged, polar or non-polar, etc.) such that the substitutions of even critical amino acids do not substantially alter activity. Conservative substitution tables providing functionally similar amino acids are well-known in the art. See, e.g., Creighton (1984) *Proteins*, W.H. Freeman and Company. In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence are also "conservatively modified variations."

The term "naturally-occurring" as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism that can be isolated from a source in nature and which has not been intentionally modified by humans in the laboratory is naturally-occurring.

The term "antibody" refers to a protein consisting of one or more polypeptides substantially encoded by immunoglobulin genes or fragments of immunoglobulin genes. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively.

A typical immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (VL) and variable heavy chain (VH) refer to these light and heavy chains respectively.

Antibodies exist as intact immunoglobulins or as a number of well-characterized fragments produced by digestion with various peptidases. Thus, for example, pepsin digests an antibody below the disulfide linkages in the hinge region to produce F(ab)₂, a dimer of Fab which itself is a light chain joined to VH-CH1 by a disulfide bond. The F(ab)₂ may be reduced under mild conditions to break the disulfide linkage in the hinge region thereby converting the (Fab)₂

dimer into an Fab' monomer. The Fab' monomer is essentially an Fab with part of the hinge region (see, *Fundamental Immunology*, W. E. Paul, ed., Raven Press, N.Y. (1993), for a more detailed description of other antibody fragments). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such Fab' fragments may be synthesized de novo either chemically or by utilizing recombinant DNA methodology. Thus, the term antibody, as used herein also includes antibody fragments either produced by the modification of whole antibodies or synthesized de novo using recombinant DNA methodologies. Preferred antibodies include single chain antibodies, more preferably single chain Fv (scFv) antibodies in which a variable heavy and a variable light chain are joined together (directly or through a peptide linker) to form a continuous polypeptide.

A single chain Fv ("scFv" or "scFv") polypeptide is a covalently linked VH::VL heterodimer which may be expressed from a nucleic acid including VH- and VL-encoding sequences either joined directly or joined by a peptide-encoding linker. Huston, et al. *Proc. Nat. Acad. Sci. USA*, 85:5879-5883 (1988). A number of structures for converting the naturally aggregated—but chemically separated light and heavy polypeptide chains from an antibody V region into an scFv molecule which will fold into a three dimensional structure substantially similar to the structure of an antigen-binding site. See, e.g. U.S. Pat. Nos. 5,091,513 and 5,132,405 and 4,956,778.

An "antigen-binding site" or "binding portion" refers to the part of an immunoglobulin molecule that participates in antigen binding. The antigen binding site is formed by amino acid residues of the N-terminal variable ("V") regions of the heavy ("H") and light ("L") chains. Three highly divergent stretches within the V regions of the heavy and light chains are referred to as "hypervariable regions" which are interposed between more conserved flanking stretches known as "framework regions" or "FRs". Thus, the term "FR" refers to amino acid sequences that are naturally found between and adjacent to hypervariable regions in immunoglobulins. In an antibody molecule, the three hypervariable regions of a light chain and the three hypervariable regions of a heavy chain are disposed relative to each other in three dimensional space to form an antigen binding "surface". This surface mediates recognition and binding of the target antigen. The three hypervariable regions of each of the heavy and light chains are referred to as "complementarity determining regions" or "CDRs" and are characterized, for example by Kabat et al. *Sequences of proteins of immunological interest*, 4th ed. U.S. Dept. Health and Human Services, Public Health Services, Bethesda, Md. (1987).

The term "antigenic determinant" refers to the particular chemical group of a molecule that confers antigenic specificity.

The term "epitope" generally refers to that portion of an antigen that interacts with an antibody. More specifically, the term epitope includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Specific binding exists when the dissociation constant for antibody binding to an antigen is $\leq 1 \mu\text{M}$, preferably $\leq 100 \text{ nM}$ and most preferably $\leq 1 \text{ nM}$. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids and typically have specific three dimensional structural characteristics, as well as specific charge characteristics.

The term "patient" includes human and veterinary subjects.

DETAILED DESCRIPTION

This invention provides a novel purified protein that the present inventors call PX3.101 which is effective in treating a variety of diseases, especially inflammatory diseases such as rheumatoid arthritis. Also provided are nucleic acids which encode PX3.101, as well as expression cassettes, expression vectors and cells containing same for use in producing PX3.101 poly-peptide and fragments thereof via recombinant methods. The present invention further provides antibodies which specifically bind to the proteins. Pharmaceutical compositions containing the proteins of the invention and methods for treating various diseases using such pharmaceutical compositions are also provided. The invention further provides methods for inhibiting the binding between chemokines and their receptors and methods for inhibiting certain enzymes associated with various inflammatory diseases.

In addition to being useful in treating various diseases such as inflammatory diseases, the protein provided by the present invention is useful in studying interactions between chemokines and receptors therefor and for kinetic and inhibition studies involving enzymes such as cyclooxygenases, phospholipases, and proteases that have been implicated in various inflammatory diseases. The nucleotide and peptide sequences provided by the present invention is also useful to generate primers and/or probes to screen for PX3.101 homologues in different species, particularly in human.

I. Proteins

In one embodiment, the present invention provides a substantially pure PX3.101 polypeptide isolated from natural sources, and/or prepared according to recombinant methods, and/or prepared by chemical synthesis, and/or using a combination of recombinant methods and chemical synthesis. PX3.101 polypeptide is exemplified by the amino acid sequence shown in FIG. 3A and SEQ ID NO:2. If isolated from natural sources, PX3.101 is preferably isolated from an insect, particularly from honey bee venom. Full-length PX3.101 has a molecular weight of approximately 7,700 daltons and has a structure characterized by five Gly-Gly-Xaa repeats (Gly=Glycine and Xaa=any amino acid; sometimes simply referred to as GGX) at the amino-terminus and a cysteine-rich motif at the carboxy terminus. As used herein the term "PX3.101" includes the full-length molecule as set forth in SEQ ID NO:2 and other polypeptides having a similar activity. The term also includes the protein lacking the signal sequence (residues 1-19 of SEQ ID NO:2). Also included, for example, are polypeptides having amino acid sequences consisting of residues 22-92 of SEQ ID NO:2, residues 24-92 of SEQ ID NO:2 and residues 26-92 of SEQ ID NO:2.

The invention also includes an isolated polypeptide having an amino acid sequence at least about 75% identical to an amino acid sequence as set forth in SEQ ID NO:2. More preferably, the polypeptide of the invention is at least 80-85% identical, still more preferably at least 90% or 95% identical to the amino acid sequence of SEQ ID NO:2. The region of similarity between PX3.101 and a polypeptide of interest typically extends over a region of at least 40 amino acids in length, more preferably over a longer region than 40 amino acids such as 50, 60, 70 or 80 amino acids, and most preferably over the full length of the polypeptide. One example of an algorithm that is useful for comparing a polypeptide to the amino acid sequence of PX3.101 is the BLASTP algorithm; suitable parameters include a word length (W) of 3, and a BLOSUM62 scoring matrix.

Besides substantially full-length polypeptides, the present invention provides for biologically active fragments of the

polypeptides. Biological activity may include effectiveness of the polypeptide in alleviating the symptoms of various inflammatory diseases (for example, rheumatoid arthritis), and/or inhibiting the binding of chemokines (e.g., IL8) with receptors therefor, and/or inhibiting enzymes associated with inflammatory diseases (such proteins include, by way of illustration and not limitation, cyclooxygenases, phospholipases, lipooxygenases, and proteases such as trypsin and cathepsin G). Other examples of significant biological activity include antibody binding (e.g., the fragment competes with a full-length PX3.101 as set forth in SEQ ID NO:2) and immunogenicity (i.e., possession of epitopes that stimulate B- or T-cell responses against the fragment).

The invention further provides a subsequence which ordinarily comprises at least 5 contiguous amino acids, typically at least 6 or 7 contiguous amino acids, more typically 8 or 9 contiguous amino acids, usually at least 10, 11 or 12 contiguous amino acids, preferably at least 13 or 14 contiguous amino acids, more preferably at least 16 contiguous amino acids, and most preferably at least 20, 40, 60 or 80 contiguous amino acids. Other examples of subsequences provided by the invention are amino acid sequences wherein 1 to 10 amino acids are removed from the N-terminal end of PX3.101 (i.e., residues 1-10 of SEQ ID NO:2). Examples of such polypeptides are listed in Table II below, wherein 2, 4 or 6 amino acids are missing from the N-terminal end of full-length PX3.101.

Polypeptides of the invention also include particular regions or domains of the amino acid sequence as set forth in SEQ ID NO:2. For example, polypeptides of the invention include the signal region (from residue 1 to 19 of SEQ ID NO:2), a region containing GGX repeats (from residue 20 to 34 of SEQ ID NO:2; also referred to as the GGX protein or peptide) and a cysteine rich region at the C-terminus (from residue 35 to 92 of SEQ ID NO:2) which is characterized by a specific cysteine pattern CXCXXG (C=Cysteine, G=Glycine, and X=any amino acid).

The polypeptides of the invention are typically encoded by nucleotide sequences that are substantially identical with the nucleotide sequence set forth in SEQ ID NO:1 and shown in FIG. 3A. The nucleotides encoding the polypeptides of the invention will also typically hybridize to the polynucleotide sequence set forth in SEQ ID NO:1.

Often the polypeptides of the invention will share at least one antigenic determinant in common with the amino acid sequence set forth in SEQ ID NO:2. The existence of such a common determinant is evidenced by cross-reactivity of the variant protein with any antibody prepared against PX3.101 polypeptide. Cross-reactivity may be tested using polyclonal sera against PX3.101, but can also be tested using one or more monoclonal antibodies against PX3.101.

The invention further includes the polypeptides described herein in which the polypeptide includes modified polypeptide backbones. Illustrative examples of such modifications include chemical derivatizations of polypeptides, such as acetylations and carboxylations. Modifications also include glycosylation modifications and processing variants of a typical polypeptide. Such processing steps specifically include enzymatic modifications, such as ubiquitination and phosphorylation. See, e.g., Hershko & Ciechanover, *Ann. Rev. Biochem.* 51:335-364 (1982).

The polypeptides provided by the invention also include isolated polypeptides comprising a signal peptide, a segment containing multiple GGX repeats (where G is glycine and N is any amino acid), and a C-terminal segment extending from the last GGX repeat to the C-terminus which contains

multiple cysteine residues. As used herein, a signal peptide or sequence is a sequence which is capable of mediating the transport of a polypeptide to the cell surface or exterior of intracellular membranes. The polypeptide is typically at least 50, 60, 70, 80 or 90 amino acids long, or any of the lengths therebetween. The polypeptide generally is no longer than 150, 140, 130, 120, 110 or 100 amino acids, or any length therebetween. The segment containing multiple GGX repeats typically contains at least 3 or 4 repeats, and in other instances contains 5 repeats, although more repeats are possible. The number of cysteines in the C-terminal segment is typically at least 5, but may be 6, 7, 8, 9, 10, 11 or 12. More cysteines than this may also be included within this segment. In one particular polypeptide, the polypeptide includes a signal sequence, a segment containing 5 GGX repeats and a C-terminal segment which includes 10 cysteines.

II. Nucleic Acids

The present invention further provides isolated and/or recombinant nucleic acids that encode the entire PX3.101 protein (SEQ ID NO:2) or subsequences thereof which have PX3.101 activity. The nucleic acids of the invention can include naturally occurring, synthetic, and intentionally manipulated polynucleotide sequences (e.g., site directed mutagenesis or use of alternate promoters for RNA transcription). The polynucleotide sequence for PX3.101 includes antisense sequences. The nucleic acids of the invention also include sequences that are degenerate as a result of the degeneracy of the genetic code.

The polynucleotide encoding PX3.101 includes the nucleotide sequence as set forth in SEQ ID NO:1 and nucleic acid sequences complementary to that sequence. Also included in the invention are subsequences of the above-described nucleic acid sequences. Such subsequences include, for example, the coding region of SEQ ID NO:1 (nucleotides 74 to 349), as well as subsequences that are at least 17, 18, 19, 20, 21, 22, 23, 24 or 25 nucleotides in length and which hybridize specifically to a nucleic acid which encodes PX3.101. Thus, the invention also includes an isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of (a) a deoxyribonucleotide sequence complementary to nucleotides 74 to 349 of SEQ ID NO:1; (b) a ribonucleotide sequence complementary to nucleotides 74 to 349 of SEQ ID NO:1; (c) a nucleotide sequence complementary to the deoxyribonucleotide sequence of (a) or to the ribonucleotide sequence of (b); (d) a nucleotide sequence of at least 23 consecutive nucleotides capable of hybridizing to nucleotides 74 to 349 of SEQ ID NO:1; and (e) a nucleotide sequence capable of hybridizing to a nucleotide sequence of (d).

The invention further provides nucleic acid molecules that include a polynucleotide sequence that encodes a polypeptide having an amino acid sequence that is substantially identical to the amino acid sequence set forth in SEQ ID NO:2. For example, the invention includes a polynucleotide sequence that encodes a polypeptide having an amino acid sequence that is at least 75% identical to the amino acid sequence as set forth in SEQ ID NO:2 over a region of at least 40 amino acids in length. More preferably, the polypeptide encoded by the nucleic acid of the invention are at least 80 to 85% identical to the amino acid sequence of SEQ ID NO:2, and still more preferably at least 90% or 95% identical to the amino acid sequence of SEQ ID NO:2 over a region of at least 40 amino acids. In some instances, the region of percent identity extends over a region of 50, 60, 70 or 80 amino acids, and more preferably over the full length of the amino acid sequence set forth in SEQ ID NO:2.

Sequence comparisons of the protein encoded by the nucleic acids of the invention can be performed visually or with a comparison algorithm. One such algorithm is the BLASTP algorithm using a wordlength (W) of 3 and the BLOSUM62 scoring matrix.

The polynucleotide sequences are typically substantially identical to a polynucleotide sequence such as residues 74 to 349 of SEQ ID NO:1. For example, the invention includes polynucleotide sequences that are at least about 75% identical to the nucleic acid SEQ ID NO:1 over a region of at least about 50 nucleotides in length. More preferably, the nucleic acids of the invention are at least 80–85% identical to the nucleic acid sequence shown in SEQ ID NO:1, and still more preferably at least 90–95% identical to the nucleic acid sequence of SEQ ID NO:1 over a region of at least 50 amino acids. In some instances, the region of percent identity extends over a longer region than 50 nucleotides, such as 75, 100, 125, 150, 175, 200, 225 or 250 nucleotides, or over the full length of the encoding region (residues 74 to 349 of SEQ ID NO:1).

To identify nucleic acids of the invention, one can employ a nucleotide sequence comparison algorithm such as are known to those of skill in the art. For example, one can use the BLASTN algorithm. Suitable parameters for use in BLASTN are wordlength (W) of 11, M=5 and N=–4. One example of a nucleic acid of the invention includes a polynucleotide sequence as set forth in SEQ ID NO:1, especially as obtained from an insect such as a bee.

Alternatively, one can identify a nucleic acid of the invention by hybridizing, under stringent conditions, the nucleic acid of interest to a nucleic acid that includes a polynucleotide sequence of SEQ ID NO:1. The invention also includes nucleic acids which encode a polypeptide which is immunologically cross reactive with PX3.101 or subsequences thereof.

Nucleic acid sequences of the present invention can be obtained by any suitable method known in the art, including, for example, 1) hybridization of genomic or cDNA libraries with probes to detect homologous nucleotide sequences; 2) antibody screening of expression libraries to detect cloned DNA fragments with shared structural features; 3) various amplification procedures such as polymerase chain reaction (PCR) using primers capable of annealing to the nucleic acid of interest; 4) direct chemical synthesis.

In one embodiment, a nucleic acid of the invention is isolated by routine cloning methods. The nucleotide sequence of a gene or cDNA encoding PX3.101 as provided herein, is used to provide probes that specifically hybridize to a PX3.101 cDNA in a cDNA library, a PX3.101 gene in a genomic DNA sample, or to a PX3.101 mRNA in a total RNA sample (e.g., in a Southern or Northern blot). Once the target nucleic acid is identified, it can be isolated according to standard methods known to those of skill in the art.

The desired nucleic acids can also be cloned using well-known amplification techniques. Examples of protocols sufficient to direct persons of skill through in vitro amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Q β -replicase amplification and other RNA polymerase mediated techniques, are found in Berger, Sambrook, and Ausubel, as well as Mullis et al. (1987) U.S. Pat. No. 4,683,202; *PCR Protocols A Guide to Methods and Applications* (Innis et al. eds) Academic Press Inc. San Diego, Calif. (1990) (Innis); Arnheim & Levinson (Oct. 1, 1990) *C&EN* 36–47; *The Journal of NIH Research* (1991) 3: 81–94; (Kwoh et al (1989) *Proc. Natl. Acad. Sci. USA* 86: 1173; Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87: 1874; Lomell et al. (1989) *J. Clin.*

Chem. 35: 1826; Landegren et al. (1988) *Science* 241: 1077–1080; Van Brunt (1990) *Biotechnology* 8: 291–294; Wu and Wallace (1989) *Gene* 4: 560; and Barringer et al. (1990) *Gene* 89: 117. Improved methods of cloning in vitro amplified nucleic acids are described in Wallace et al., U.S. Pat. No. 5,426,039. Suitable primers for use in the amplification of the nucleic acids of the invention include, for example, forward primer ASEQ10: 5' AAGGATCCA-CAGTGAACGTAAGTTC 3' (SEQ ID NO:3), forward primer ASEQ11: 5' AAGGATCCGGAGGATTTGGAG-GATTTGGAGGATTTGGAGGACTTGGAGGACGTGG 3' (SEQ ID NO:4), reverse primer ASEQ13: 5'ACT-GATAAAATAATAAC 3' (SEQ ID NO:5), reverse primer ASEQ14: 5' ATGAATGATAAAATAC 3' (SEQ ID NO:6), reverse primer ASEQ15: 5'TTATAAAAGTCATCCGC 3'(SEQ ID NO:7).

As an alternative to cloning a nucleic acid, a suitable nucleic acid can be chemically synthesized. Direct chemical synthesis methods include, for example, the phosphotriester method of Narang et al. (1979) *Meth. Enzymol.* 68: 90–99; the phosphodiester method of Brown et al. (1979) *Meth. Enzymol.* 68: 109–151; the diethylphosphoramidite method of Beaucage et al. (1981) *Tetra. Lett.*, 22: 1859–1862; and the solid support method of U.S. Pat. No. 4,458,066. Chemical synthesis produces a single stranded oligonucleotide. This can be converted into double stranded DNA by hybridization with a complementary sequence, or by polymerization with a DNA polymerase using the single strand as a template. One of skill would recognize that while chemical synthesis of DNA is often limited to sequences of about 100 bases, longer sequences may be obtained by the ligation of shorter sequences. Alternatively, subsequences may be cloned and the appropriate subsequences cleaved using appropriate restriction enzymes. The fragments can then be ligated to produce the desired DNA sequence.

In some embodiments, it may be desirable to modify the nucleic acids of the invention. One of skill will recognize many ways of generating alterations in a given nucleic acid construct. Such well-known methods include site-directed mutagenesis, PCR amplification using degenerate oligonucleotides, exposure of cells containing the nucleic acid to mutagenic agents or radiation, chemical synthesis of a desired oligonucleotide (e.g., in conjunction with ligation and/or cloning to generate large nucleic acids) and other well-known techniques. See, e.g., Gilman and Smith (1979) *Gene* 8:81–97, Roberts et al. (1987) *Nature* 328: 731–734.

III. Methods for Preparing or Isolating Protein

A. Recombinant Technologies

1. General

The nucleotide and amino acid sequences of PX3.101 as shown in SEQ ID NO:1 and SEQ ID NO:2, respectively, and corresponding sequences for other variants as described above, allow for production of full-length PX3.101 polypeptide and fragments thereof. Such polypeptides can be produced in prokaryotic or eukaryotic host cells by expression of polynucleotides encoding PX3.101 or fragments thereof. The cloned DNA sequences are expressed in hosts after the sequences have been operably linked to an expression control sequence in an expression vector. Expression vectors are typically replicable in the host organisms either as episomes or as an integral part of the host chromosomal DNA. Commonly, expression vectors contain selection markers, e.g., tetracycline resistance or hygromycin resistance, to permit detection and/or selection of those cells transformed with the desired DNA sequences (see, e.g., U.S. Pat. No. 4,704,362).

2. Expression Cassettes and Host Cells for Expressing Polypeptides

Typically, the polynucleotide that encodes a polypeptide of the invention is placed under the control of a promoter that is functional in the desired host cell to produce relatively large quantities of a polypeptide of the invention. An extremely wide variety of promoters are well-known, and can be used in the expression vectors of the invention, depending on the particular application. Ordinarily, the promoter selected depends upon the cell in which the promoter is to be active. Other expression control sequences such as ribosome binding sites, transcription termination sites and the like are also optionally included. Constructs that include one or more of these control sequences are termed "expression cassettes." Accordingly, the invention provides expression cassettes into which the nucleic acids that encode the polypeptides described herein are incorporated for high level expression in a desired host cell.

In a preferred embodiment, the expression cassettes are useful for expression of the polypeptides of the invention in prokaryotic host cells. Commonly used prokaryotic control sequences, which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the beta-lactamase (penicillinase) and lactose (lac) promoter systems (Change et al. (1977) *Nature* 198: 1056), the tryptophan (trp) promoter system (Goeddel et al. (1980) *Nucleic Acids Res.* 8: 4057), the tac promoter (DeBoer et al. (1983) *Proc. Natl. Acad. Sci. U.S.A.* 80:21-25); and the lambda-derived P_L promoter and N-gene ribosome binding site (Shimatake et al. (1981) *Nature* 292: 128). The particular promoter system is not critical to the invention, any available promoter that functions in prokaryotes can be used.

For expression of polypeptides in prokaryotic cells other than *E. coli*, a promoter that functions in the particular prokaryotic species is required. Such promoters can be obtained from genes that have been cloned from the species, or heterologous promoters can be used. For example, the hybrid trp-lac promoter functions in *Bacillus* in addition to *E. coli*.

For expression of the polypeptides in yeast, convenient promoters include GAL1-10 (Johnson and Davies (1984) *Mol. Cell. Biol.* 4:1440-1448) ADH2 (Russell et al. (1983) *J. Biol. Chem.* 258:2674-2682), PHO5 (*EMBO J.* (1982) 6:675-680), and MF α (Herskowitz and Oshima (1982) in *The Molecular Biology of the Yeast Saccharomyces* (eds. Strathern, Jones, and Broach) Cold Spring Harbor Lab., Cold Spring Harbor, N.Y., pp. 181-209). Another suitable promoter for use in yeast is the ADH2/GAPDH hybrid promoter as described in Cousens et al., *Gene* 61:265-275 (1987). Other promoters suitable for use in eukaryotic host cells are well-known to those of skill in the art.

For expression of the polypeptides in mammalian cells, convenient promoters include CMV promoter (Miller, et al., *BioTechniques* 7:980), SV40 promoter (de la Luma, et al., (1998) *Gene* 62:121), RSV promoter (Yates, et al, (1985) *Nature* 313:812), MMTV promoter (Lee, et al., (1981) *Nature* 294:228).

For expression of the polypeptides in insect cells, the convenient promoter is from the baculovirus *Autographa Californica* nuclear polyhedrosis virus (NcMNPV) (Kitts, et al., (1993) *Nucleic Acids Research* 18:5667).

Either constitutive or regulated promoters can be used in the present invention. Regulated promoters can be advantageous because the host cells can be grown to high densities before expression of the polypeptides is induced. High level

expression of heterologous proteins slows cell growth in some situations. An inducible promoter is a promoter that directs expression of a gene where the level of expression is alterable by environmental or developmental factors such as, for example, temperature, pH, anaerobic or aerobic conditions, light, transcription factors and chemicals. Such promoters are referred to herein as "inducible" promoters, and allow one to control the timing of expression of the polypeptide. For *E. coli* and other bacterial host cells, inducible promoters are known to those of skill in the art. These include, for example, the lac promoter, the bacteriophage lambda P_L promoter, the hybrid trp-lac promoter (Amann et al. (1983) *Gene* 25: 167; de Boer et al. (1983) *Proc. Nat'l. Acad. Sci. USA* 80: 21), and the bacteriophage T7 promoter (Studier et al. (1986) *J. Mol. Biol.*; Tabor et al. (1985) *Proc. Nat'l. Acad. Sci. USA* 82: 1074-8). These promoters and their use are discussed in Sambrook et al., supra. A particularly preferred inducible promoter for expression in prokaryotes is a dual promoter that includes a tac promoter component linked to a promoter component obtained from a gene or genes that encode enzymes involved in galactose metabolism (e.g., a promoter from a UDP galactose 4-epimerase gene (galI)). The dual tac-gal promoter, which is described in PCT Patent Application Publ. No. WO98/20111, provides a level of expression that is greater than that provided by either promoter alone.

Inducible promoters for other organisms are also well-known to those of skill in the art. These include, for example, the arabinose promoter, the lacZ promoter, the metallothionein promoter, and the heat shock promoter, as well as many others.

A ribosome binding site (RBS) is conveniently included in the expression cassettes of the invention that are intended for use in prokaryotic host cells. An RBS in *E. coli*, for example, consists of a nucleotide sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon (Shine and Dalgarno (1975) *Nature* 254: 34; Steitz, *In Biological regulation and development: Gene expression* (ed. R. F. Goldberger), vol. 1, p. 349, 1979, Plenum Publishing, NY).

Selectable markers are often incorporated into the expression vectors used to express the polynucleotides of the invention. These genes can encode a gene product, such as a protein, necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics or other toxins, such as ampicillin, neomycin, kanamycin, chloramphenicol, or tetracycline. Alternatively, selectable markers may encode proteins that complement auxotrophic deficiencies or supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for *Bacilli*. Often, the vector will have one selectable marker that is functional in, e.g., *E. coli*, or other cells in which the vector is replicated prior to being introduced into the host cell. A number of selectable markers are known to those of skill in the art and are described for instance in Sambrook et al., supra. A preferred selectable marker for use in bacterial cells is a kanamycin resistance marker (Vieira and Messing, *Gene* 19: 259 (1982)). Use of kanamycin selection is advantageous over, for example, ampicillin selection because ampicillin is quickly degraded by β -lactamase in culture medium, thus removing selective pressure and allowing the culture to become overgrown with cells that do not contain the vector.

Construction of suitable vectors containing one or more of the above listed components employs standard ligation

techniques as described in the references cited above. Isolated plasmids or DNA fragments are cleaved, tailored, and re-ligated in the form desired to generate the plasmids required. To confirm correct sequences in plasmids constructed, the plasmids can be analyzed by standard techniques such as by restriction endonuclease digestion, and/or sequencing according to known methods. A wide variety of cloning and in vitro amplification methods suitable for the construction of recombinant nucleic acids are well-known to persons of skill. Examples of these techniques and instructions sufficient to direct persons of skill through many cloning exercises are found in Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology*, Volume 152, Academic Press, Inc., San Diego, Calif. (Berger); and Current Protocols in Molecular Biology, F. M. Ausubel et al., eds., *Current Protocols*, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1998 Supplement) (Ausubel).

A variety of common vectors suitable for use as starting materials for constructing the expression vectors of the invention are well-known in the art. For cloning in bacteria, common vectors include pBR322 derived vectors such as pBLUESCRIPT™, pUC18/19, and λ-phage derived vectors. In yeast, vectors which can be used include Yeast Integrating plasmids (e.g., YIp5) and Yeast Replicating plasmids (the YRp series plasmids) pYES series and pGPD-2 for example. Expression in mammalian cells can be achieved, for example, using a variety of commonly available plasmids, including pSV2, pBC12BI, and p91023, pCDNA series, pCMV1, pMAMneo, as well as lytic virus vectors (e.g., vaccinia virus, adeno virus), episomal virus vectors (e.g., bovine papillomavirus), and retroviral vectors (e.g., murine retroviruses). Expression in insect cells can be achieved using a variety of baculovirus vectors, including pFastBac1, pFastBacHT series, pBluesBac4.5, pBluesBacHis series, pMelBac series, and pVL1392/1393, for example.

Translational coupling can be used to enhance expression. The strategy uses a short upstream open reading frame derived from a highly expressed gene native to the translational system, which is placed downstream of the promoter, and a ribosome binding site followed after a few amino acid codons by a termination codon. Just prior to the termination codon is a second ribosome binding site, and following the termination codon is a start codon for the initiation of translation. The system dissolves secondary structure in the RNA, allowing for the efficient initiation of translation. See, Squires et al. (1988) *J. Biol. Chem.* 263: 16297-16302.

Polypeptides of the invention can be expressed in a variety of host cells, including *E. coli*, other bacterial hosts, yeast, and various higher eukaryotic cells such as the COS, CHO and HeLa cells lines and myeloma cell lines. The host cells can be mammalian cells, plant cells, insect cells or microorganisms, such as, for example, yeast cells, bacterial cells, or fungal cells. Examples of suitable host cells include *Azotobacter* sp. (e.g., *A. vinelandii*), *Pseudomonas* sp., *Rhizobium* sp., *Erwinia* sp., *Escherichia* sp. (e.g., *E. coli*), *Bacillus*, *Pseudomonas*, *Proteus*, *Salmonella*, *Serratia*, *Shigella*, *Rhizobia*, *Vitreoscilla*, *Paracoccus* and *Klebsiella* sp., among many others. The cells can be of any of several genera, including *Saccharomyces* (e.g., *S. cerevisiae*), *Candida* (e.g., *C. utilis*, *C. parapsilosis*, *C. krusei*, *C. versatilis*, *C. lipolytica*, *C. zeylanoides*, *C. guilliermondii*, *C. albicans*, and *C. humicola*), *Pichia* (e.g., *P. farinosa* and *P. ohmeri*), *Torulopsis* (e.g., *T. candida*, *T. sphaerica*, *T. xylinus*, *T. famata*, and *T. versatilis*), *Debaryomyces* (e.g., *D. subglobosus*, *D. cantarellii*, *D. globosus*, *D. hansenii*, and *D. japonicus*), *Zygosaccharomyces* (e.g., *Z. rouxii* and *Z.*

bailii), *Kluyveromyces* (e.g., *K. marxianus*), *Hansenula* (e.g., *H. anomala* and *H. jadinii*), and *Brettanomyces* (e.g., *B. lambicus* and *B. anomalus*). Examples of useful bacteria include, but are not limited to, *Escherichia*, *Enterobacter*, *Azotobacter*, *Erwinia*, *Klebsiella*. The commonly used insect cells to produce recombinant proteins are Sf9 cells (derived from *Spodoptera frugiperda* ovarian cells) and High Five cells (derived from *Trichoplusia ni* egg cell homogenates; commercially available from Invitrogen).

The expression vectors of the invention can be transferred into the chosen host cell by well-known methods such as calcium chloride transformation for *E. coli* and calcium phosphate treatment or electroporation for mammalian cells. Cells transformed by the plasmids can be selected by resistance to antibiotics conferred by genes contained on the plasmids, such as the amp, gpt, neo and hyg genes.

Once expressed, the recombinant polypeptides can be purified according to standard procedures of the art, including ammonium sulfate precipitation, affinity columns, ion exchange and/or size exclusivity chromatography, gel electrophoresis and the like (see, generally, R. Scopes, *Protein Purification*, Springer-Verlag, N.Y. (1982), Deutscher, *Methods in Enzymology Vol 182: Guide to Protein Purification*, Academic Press, Inc. N.Y. (1990)). Substantially pure compositions of at least about 90 to 95% homogeneity are preferred, and 98 to 99% or more homogeneity are most preferred. Once purified, partially or to homogeneity as desired, the polypeptides may then be used (e.g., treatment of inflammatory diseases in pre-clinical or clinical studies).

To facilitate purification of the polypeptides of the invention, the nucleic acids that encode the polypeptides can also include a coding sequence for an epitope or "tag" for which an affinity binding reagent is available. Examples of suitable epitopes include the myc and V-5 reporter genes; expression vectors useful for recombinant production of polypeptides having these epitopes are commercially available (e.g., Invitrogen (Carlsbad Calif.) vectors pcDNA3.1/Myc-His and pcDNA3.1/V5-His are suitable for expression in mammalian cells; Invitrogen (Carlsbad, Calif.) vectors pBlueBacHis and Gibco (Gaithersburg, Md.) vectors pFastBacHT are suitable for expression in insect cells). Additional expression vectors suitable for attaching a tag to the proteins of the invention, and corresponding detection systems are known to those of skill in the art, and several are commercially available (e.g., FLAG" (Kodak, Rochester N.Y.). Another example of a suitable tag is a polyhistidine sequence, which is capable of binding to metal chelate affinity ligands. Typically, six adjacent histidines are used, although one can use more or less than six. Suitable metal chelate affinity ligands that can serve as the binding moiety for a polyhistidine tag include nitrilo-tri-acetic acid (NTA) (Hochuli, E. (1990) "Purification of recombinant proteins with metal chelating adsorbents" In *Genetic Engineering: Principles and Methods*, J. K. Setlow, Ed., Plenum Press, NY; commercially available from Qiagen (Santa Clarita, Calif.)).

Other haptens that are suitable for use as tags are known to those of skill in the art and are described, for example, in the *Handbook of Fluorescent Probes and Research Chemicals* (6th Ed., Molecular Probes, Inc., Eugene Oreg.). For example, dinitrophenol (DNP), digoxigenin, barbiturates (see, e.g., U.S. Pat. No. 5,414,085), and several types of fluorophores are useful as haptens, as are derivatives of these compounds. Kits are commercially available for linking haptens and other moieties to proteins and other molecules. For example, where the hapten includes a thiol, a heterobi-

functional linker such as SMCC can be used to attach the tag to lysine residues present on the capture reagent.

B. Naturally-Occurring Polypeptides

Naturally occurring polypeptides of the invention, including full-length PX3.101 and fragments thereof can be isolated using conventional techniques such as affinity chromatography. For example, polyclonal or monoclonal antibodies are raised against previously-purified PX3.101 or fragments thereof and attached to a suitable affinity column by well-known techniques. See, e.g., Hudson & Hay, *Practical Immunology* (Blackwell Scientific Publications, Oxford, UK, 1980), Chapter 8 (incorporated herein by reference in its entirety for all purposes). Peptide fragments are generated from intact PX3.101 by chemical or enzymatic cleavage methods which are known to those with skill in the art. Example II also sets forth a method for purifying PX3.101.

C. Other Methods

Alternatively, the polypeptides of the invention can be synthesized by chemical methods or produced by in vitro translation systems using a polynucleotide template to direct translation. Methods for chemical synthesis of polypeptides and in vitro translation are well-known in the art, and are described further by Berger & Kimmel, *Methods in Enzymology*, Volume 152, *Guide to Molecular Cloning Techniques*, Academic Press, Inc., San Diego, Calif., 1987 (incorporated herein by reference in its entirety for all purposes).

IV. Antibodies

In another embodiment of the invention, antibodies that are immunoreactive with PX3.101 polypeptide or fragments thereof are provided. The antibodies may be polyclonal antibodies, distinct monoclonal antibodies or pooled monoclonal antibodies with different epitopic specificities. Monoclonal antibodies are made from antigen-containing fragments of the protein by methods that are well-known in the art (see, e.g., Kohler, et al. *Nature*, 256:495, (1975); and Harlow & Lane, *Antibodies, A Laboratory Manual* (C.S.H.P., NY, 1988), both of which are incorporated herein by reference in their entirety for all purposes).

A. Production of Antibodies

Antibodies that bind to PX3.101 polypeptide or other polypeptides of the invention can be prepared using intact polypeptide or fragments containing small peptides of interest as the immunizing antigen. For example, it may be desirable to produce antibodies that specifically bind to the N- or C-terminal domains of PX3.101. The polypeptide used to immunize an animal can be from natural sources, derived from translated cDNA, or prepared by chemical synthesis and can be conjugated with a carrier protein, if desired. Commonly used carriers include keyhole limpet hemocyanin (KLH), thyroglobulin, bovine serum albumin (BSA), and tetanus toxoid. The coupled peptide is then used to immunize the animal (e.g., a mouse, a rat, or a rabbit).

Techniques for generation of human monoclonal antibodies have also been described but are generally more onerous than murine techniques and not applicable to all antigens. See, e.g., Larrick et al., U.S. Pat. No. 5,001,065, for review (incorporated by reference for all purposes). An alternative approach is the generation of humanized antibodies by linking the complementarity-determining regions or CDR regions (see, e.g., Kabat et al., "Sequences of Proteins of Immunological Interest," U.S. Dept. of Health and Human Services, (1987); and Chothia et al., *J. Mol. Biol.* 196:901-917 (1987)) of non-human antibodies to human constant regions by recombinant DNA techniques. See Queen et al., *Proc. Natl. Acad. Sci. USA* 86:10029-10033

(1989) and WO 90/07861, (incorporated by reference for all purposes). Alternatively, one may isolate DNA sequences which encode a human monoclonal antibody or a binding fragment thereof by screening a DNA library from human B cells according to the general protocol outlined by Huse et al., *Science* 246:1275-1281 (1989) and then cloning and amplifying the sequences which encode the antibody (or binding fragment) of the desired specificity. The protocol described by Huse is rendered more efficient in combination with phage display technology. See, e.g., Dower et al., WO 91/17271 and McCafferty et al., WO 92/01047 (each of which is incorporated by reference for all purposes). Phage display technology can also be used to mutagenize CDR regions of antibodies previously shown to have affinity for the peptides of the present invention. Antibodies having improved binding affinity are selected.

The antibodies can be further purified, for example, by binding to and elution from a support to which the polypeptide or a peptide to which the antibodies were raised is bound. A variety of other techniques known in the art can also be used to purify polyclonal or monoclonal antibodies (see, e.g., Coligan, et al., Unit 9, *Current Protocols in Immunology*, Wiley Interscience, (1994), incorporated herein by reference in its entirety).

It is also possible to use the anti-idiotypic technology to produce monoclonal antibodies which mimic an epitope. For example, an anti-idiotypic monoclonal antibody made to a first monoclonal antibody will have a binding domain in the hypervariable region which is the "image" of the epitope bound by the first monoclonal antibody.

B. Use of Antibodies

The antibodies of the invention are useful, for example, in screening cDNA expression libraries and for identifying clones containing cDNA inserts which encode structurally-related, immunocrossreactive proteins. See, for example, Aruffo & Seed, *Proc. Natl. Acad. Sci. USA* 84:8573-8577 (1977) (incorporated herein by reference in its entirety for all purposes). Antibodies are also useful to identify and/or purify immunocrossreactive proteins that are structurally related to native PX3.101 or to fragments thereof used to generate the antibody.

V. Therapeutic Methods and Compositions

A. General

The present invention further provides pharmaceutical compositions comprised of the polypeptides of the present invention, including full-length PX3.101 and fragments thereof. As explained more fully below in Example V, the pharmaceutical compositions of the invention are useful in treating a variety of diseases in both human and veterinary subjects. Diseases which can be treated with certain pharmaceutical compositions of the inventions include a variety of inflammatory diseases and autoimmune diseases, (e.g., rheumatoid arthritis, multiple sclerosis, psoriasis, systemic lupus erythematosus (SLE), Crohn's disease, scleroderma), metastatic cancers, and diseases associated with imbalances in chemokine (e.g., IL-8, IL-10, IL-1, and TNF- α) production such as Alzheimer disease. Some compositions can also be used to treat pain, i.e., the composition can be used as an analgesic.

Pharmaceutical compositions of the invention are suitable for use in a variety of drug delivery systems. Suitable formulations for use in the present invention are found in *Remington's Pharmaceutical Sciences*, Mace Publishing Company, Philadelphia, Pa., 17th ed. (1985). For a brief review of methods for drug delivery, see, Langer, *Science* 249:1527-1533 (1990).

B. Composition and Delivery

The pharmaceutical compositions used for prophylactic or therapeutic treatment comprise an active therapeutic agent, for example, a PX3.101 protein or fragments thereof, a PX3.101 receptor or fragments thereof, and antibodies and 5 idiotypic antibodies thereto, and a variety of other components. Various subsequences of full-length PX3.101 can be used. For example, the polypeptide composition used in the animal studies described in Example V included peptides having amino acid sequences consisting of residues 20-92, 22-92, 24-92 and 26-92 of SEQ ID NO:2.

In some instances, the efficacy of treatment may be enhanced by using the pharmaceutical compositions of the present invention with other complementary compounds that are known to be effective in the treatment of various diseases, especially inflammatory diseases. For example, the pharmaceutical compositions of the invention may also include a compound effective in treatment of inflammatory diseases, such as rheumatoid arthritis for example. Such compounds include ENBREL (manufactured by Immunex) and INDOMETHACIN (manufactured by Merck), METH- 15 OTREXATE (manufactured by Mylan and Roxane Laboratories, Inc.), CELEBREX (manufactured by Mosanto), VIOXX (manufactured by Merck), or CYCLOSPORINE (manufactured by Novartis). Other compounds that can be used to treat inflammatory diseases and which can be combined with certain compositions of the invention can be found in the Physician's Desk Reference (1998), which is incorporated herein by reference in its entirety. The complementary compounds used in combina- 20 tion with PX3.101, fragments thereof and/or antibodies thereto, typically have a different mode of action than PX3.101 or fragments thereof and/or differ with respect to the time period during which they are therapeutically effective. Thus, for example, a pharmaceutical composition of the invention includes a therapeutically effective amount of PX3.101 (or an active fragment thereof) in combination with ENBREL since early studies indicate that the two com- 25 pounds appear to have different mechanisms of action and different time periods during which the therapeutic effect is maintained once treatment is stopped.

The compositions may also include, depending on the formulation desired, pharmaceutically-acceptable, non-toxic carriers of diluents, which are defined as vehicles commonly used to formulate pharmaceutical compositions for animal or human administration. The diluent is selected so as not to affect the biological activity of the combination. Examples of such diluents are distilled water, buffered water, physi- 30 ological saline, PBS, Ringer's solution, dextrose solution, and Hank's solution. In addition, the pharmaceutical composition or formulation may also include other carriers, adjuvants, or non-toxic, nontherapeutic, nonimmunogenic stabilizers, excipients and the like. The compositions may also include additional substances to approximate physi- 35 ological conditions, such as pH adjusting and buffering agents, toxicity adjusting agents, wetting agents, detergents and the like.

The composition may also include any of a variety of stabilizing agents, such as an antioxidant for example. Moreover, the polypeptides may be complexed with various well-known compounds that enhance the in vivo stability of the polypeptide, or otherwise enhance its pharmacological properties (e.g., increase the half-life of the polypeptide, reduce its toxicity, enhance solubility or uptake). Examples of such modifications or complexing agents include the production of sulfate, gluconate, citrate, phosphate and the like. The polypeptides of the composition may also be 40

complexed with molecules that enhance their in vivo attributes. Such molecules include, for example, carbohydrates, polyamines, amino acids, other peptides, ions (e.g., sodium, potassium, calcium, magnesium, manganese), and lipids.

Further guidance regarding formulations that are suitable for various types of administration can be found in *Remington's Pharmaceutical Sciences*, Mace Publishing Company, Philadelphia, Pa., 17th ed. (1985). For a brief review of methods for drug delivery, see, Langer, *Science* 249:1527-1533 (1990).

The compositions containing the polypeptides can be administered for prophylactic and/or therapeutic treatments. The polypeptide in the pharmaceutical composition typi- 15 cally is present in a therapeutic amount, which is an amount sufficient to remedy a disease state or symptoms, particularly symptoms associated with inflammation, or otherwise prevent, hinder, retard, or reverse the progression of disease or any other undesirable symptoms in any way whatsoever. The concentration of the polypeptide in the pharmaceutical composition can vary widely, i.e., from less than about 0.1% by weight, usually being at least about 1% by weight, to as much as 20% by weight or more.

In therapeutic applications, compositions are adminis- 25 tered to a patient already suffering from a disease, as just described, in an amount sufficient to cure or at least partially arrest the symptoms of the disease and its complications. An appropriate dosage of the pharmaceutical composition or polypeptide of the invention is readily determined according to any one of several well-established protocols. For example, animal studies (e.g., mice, rats) are commonly used to determine the maximal tolerable dose of the bioactive agent per kilogram of weight. In general, at least one of the animal species tested is mammalian. The results from the animal studies can be extrapolated to determine doses for use in other species, such as humans for example.

What constitutes an effective dose also depends on the nature and severity of the disease or condition, and on the general state of the patient's health, but will generally range from about 1 to 500 mg of purified protein per kilogram of body weight, with dosages of from about 5 to 100 mg per kilogram being more commonly employed.

In prophylactic applications, compositions containing the compounds of the invention are administered to a patient susceptible to or otherwise at risk of a particular disease or infection. Such an amount is defined to be a "prophylactically effective" amount or dose. In this use, the precise amounts again depends on the patient's state of health and weight. Typically, the dose ranges from about 1 to 500 mg of purified protein per kilogram of body weight, with dosages of from about 5 to 100 mg per kilogram being more commonly utilized.

The pharmaceutical compositions described herein can be administered in a variety of different ways. Examples include administering a composition containing a pharmaceutically acceptable carrier via oral, intranasal, rectal, topical, intraperitoneal, intravenous, intramuscular, subcutaneous, subdermal, transdermal, intrathecal, and intracranial methods.

For oral administration, the active ingredient can be administered in solid dosage forms, such as capsules, tablets, and powders, or in liquid dosage forms, such as elixirs, syrups, and suspensions. The active component(s) can be encapsulated in gelatin capsules together with inactive ingredients and powdered carriers, such as glucose, lactose, sucrose, mannitol, starch, cellulose or cellulose derivatives, magnesium stearate, stearic acid, sodium 65

saccharin, talcum, magnesium carbonate. Examples of additional inactive ingredients that may be added to provide desirable color, taste, stability, buffering capacity, dispersion or other known desirable features are red iron oxide, silica gel, sodium lauryl sulfate, titanium dioxide, and edible white ink. Similar diluents can be used to make compressed tablets. Both tablets and capsules can be manufactured as sustained release products to provide for continuous release of medication over a period of hours. Compressed tablets can be sugar coated or film coated to mask any unpleasant taste and protect the tablet from the atmosphere, or enteric-coated for selective disintegration in the gastrointestinal tract. Liquid dosage forms for oral administration can contain coloring and flavoring to increase patient acceptance.

If desired, it is possible to formulate solid or liquid formulations in an enteric-coated or otherwise protected form. In the case of liquid formulations, the formulation can be mixed or simply coadministered with a protectant, such as a liquid mixture of medium chain triglycerides, or the formulation can be filled into enteric capsules (e.g., of soft or hard gelatin, which are themselves optionally additionally enteric coated). Alternatively, solid formulations comprising the polypeptide can be coated with enteric materials to form tablets. The thickness of enteric coating on tablets or capsules can vary. Typical thickness range from 0.5 to 4 microns in thickness. The enteric coating may comprise any of the enteric materials conventionally utilized in orally administrable pharmaceutical formulations. Suitable enteric coating materials are known, for example, from *Remington's Pharmaceutical Sciences*, Mace Publishing Company, Philadelphia, 17th ed. (1985); and *Hagers Handbuch der Pharmazeutischen Praxis*, Springer Verlag, 4th ed., Vol. 7a (1971).

Another delivery option involves loading the composition into lipid-associated structures (e.g., liposomes, or other lipidic complexes) which may enhance the pharmaceutical characteristics of the polypeptide component of the composition. The complex containing the composition may subsequently be targeted to specific target cells by the incorporation of appropriate targeting molecules (e.g., specific antibodies or receptors). It is also possible to directly complex the polypeptide with a targeting agent.

Compositions prepared for intravenous administration typically contain 100 to 500 ml of sterile 0.9% NaCl or 5% glucose optionally supplemented with a 20% albumin solution and 100 to 500 mg of a polypeptide of the invention. A typical pharmaceutical composition for intramuscular injection would be made up to contain, for example, 1 ml of sterile buffered water and 1 to 10 mg of the purified polypeptide of the invention. Methods for preparing parenterally administrable compositions are well-known in the art and described in more detail in various sources, including, for example, *Remington's Pharmaceutical Science*, Mack Publishing, Philadelphia, Pa., 17th ed., (1985).

Particularly when the compositions are to be used in vivo, the components used to formulate the pharmaceutical compositions of the present invention are preferably of high purity and are substantially free of potentially harmful contaminants (e.g., at least National Food (NF) grade, generally at least analytical grade, and more typically at least pharmaceutical grade). Moreover, compositions intended for in vivo use are usually sterile. To the extent that a given compound must be synthesized prior to use, the resulting product is typically substantially free of any potentially toxic agents, particularly any endotoxins, which may be present during the synthesis or purification process. Compositions

for parental administration are also sterile, substantially isotonic and made under GMP conditions.

V. Uses

The pharmaceutical compositions of the present invention can be used to treat a variety of diseases. For example, the pharmaceutical compositions can be used in treating various inflammatory diseases. As described in more detail in Example V, certain compositions of the invention have been shown to be effective in treating rheumatoid arthritis in animal model studies. In particular, PX3.101 inhibits several enzymes that are involved in the pathogenesis of rheumatoid arthritis such as cyclooxygenases, phospholipases, lipoxigenases, and various proteases. PX3.101 polypeptide has also been shown to inhibit interaction between cytokines and their receptors (see Example VI), such as IL-8/CXCR2 interaction for example. IL-8 is a major chemokine that regulates the inflammatory process (see e.g., Harada, et al., (1994) *J. Leukoc. Biol.* 56:559). There is also evidence that links IL-8 to tumor angiogenesis and tumor metastasis (Koch, et al, *Science* 258:1798). Thus some polypeptides of the invention can be used in treating cancer, autoimmune diseases, and/or other inflammatory diseases associated with chemokine imbalances, especially diseases correlated with IL-8 such as Alzheimer disease.

PX3.101 and other polypeptides of the invention also find use in inhibition and kinetic investigations. For example, PX3.101 can be used in studies into the interaction between chemokines and receptors therefor, for example the interaction between IL-8 and CXCR2 and CXCR1. Methods involving inhibition of chemokines generally involve allowing a quantity of the chemokine and receptor to admix with a polypeptide of the invention. More specifically, the method involves adding a polypeptide of the invention to a sample containing the chemokine and receptor and preferably mixing the resulting mixture. Additions may be made to an in vitro solution or directly into a patient.

PX3.101 is also useful in studies or treatments involving inhibition of various enzymes such as cyclooxygenases (for example, COX1 and COX2), phospholipases (for example, phospholipase A2 and phospholipase C), lipoxigenase, and various proteases (for example, trypsin and cathepsin G). Such methods generally involve allowing an enzyme, particularly those involved in inflammatory processes, to admix with a polypeptide of the invention. Typically, a quantity of polypeptide is added to a sample containing the enzyme of interest. Here, too, additions may be made into an in vitro solution or directly into a patient.

Based upon the foregoing activities, it is also expected that certain polypeptides of the invention can be useful in treating blood coagulation diseases, accelerating wound healing, UV-light protection, reducing various aging phenomenon and as a pain analgesic.

By using PX3.101 to study the interaction between chemokines and their receptors, or the direct interaction between PX3.101 and a cognate receptor (e.g., the amino acid sequence which binds to a receptor), small molecules which mimic this interaction can be developed, thus enabling the small molecule to be used to obtain a therapeutic effect similar to that obtained using PX3.101.

The nucleotide and peptide sequences of PX3.101 is also useful to generate primers and/or probes to screen for PX3.101 homologues in different species, particularly in human. Human homologues of PX3.101 can be directly used as a therapeutic material or as a target to screen drug candidates for various human diseases.

The following examples are offered to further illustrate specific aspects of the present invention and are not to be interpreted so as to limit the scope of the present invention.

EXAMPLE I

General Methods and Materials

A. Electrophoresis and Western Blotting

SDS polyacrylamide gel electrophoresis (SDS-PAGE) was carried out according to standard methods (Sambrook, et al, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, 1989). Purified PX3.101 protein was electrophoresed in 10–20% SDS-polyacrylamide gradients gel (BioRad, Richmond, Calif.) and then transferred onto nitrocellulose membrane (Schleicher & Schull, Keene, N.H.). The sera collected from animals treated with PBS or PX3.101 (200 µg/kg) was diluted 1:30 in blocking solution (PBS, 0.2% Tween-20, 5% dry milk). The blot was divided using a mini-protein II multi-screening apparatus and probed with diluted sera. Horseradish Peroxidase (HRP)-conjugated goat anti-mouse IgG at 1:10,000 dilution was used as the secondary antibodies. Signals were visualized using an ECL (enhanced chemi-luminescence) system.

B. Mass Spectral Analysis

Mass spectral analysis was carried out by Heck Facility at Yale University according to the protocols outlined on their website (<http://www.info.med.yale.edu/wmkeck>). Matrix assisted laser desorption ionization mass spectrometry (MALDI-MS) was used to determine the molecular weights of the peptide fragments from the trypsin digest of PX3.101, and both MALDI-MS and electrospray ionization mass spectrometry (ESMS) were used to determine the molecular weights of the purified PX3.101 mutual derivatives. MALDI-MS was carried out on a research grade, VG Tofspec SE instrument equipped with delayed extraction and a reflectron. ESMS was carried out on a Micromass Q-Tof mass spectrometer.

C. Reverse Phase (RP) HPLC

Reverse phase HPLC (RP-HPLC) was performed using a Varian Dynamax Model SD-200 solvent delivery module and a UV detector (Dynamax Absorbance Detector Model UV-C). Data acquisition was achieved by using Varian Dynamax Method (Version 1.4.6) software. A Varian Microsorb-MV C-18 reverse phase column (0.46 cm×25 cm) was typically used for analytical analyses, while a semi-prep C-18 reverse phase column (1.0 cm×25 cm), Varian Dynamax 300 A, was used mainly for purification purposes. Buffers used for elution were Buffer A (water with 0.1% trifluoroacetic acid) and Buffer B (acetonitrile with 0.1% trifluoroacetic acid). Elution was achieved using a discontinuous linear gradient formed from buffers A and B. Columns were run at room temperature. (Flow rate: 1 ml/min for analytical HPLC; 4 ml/min for semi-prep). Peak fractions were collected by hand.

D. HPLC—Ion Exchange Chromatography

HPLC-ion exchange chromatography was also used for the purification of PX3.101. The same solvent delivery system described above was used. A Varian Hydropore strong cation exchange column (0.46 cm×10 cm) was used. Buffers used for elution were Buffer A (0.1 M ammonium formate in water, pH 5.8) and Buffer B (1 M ammonium formate, pH 6.7). Elution was achieved using a discontinuous linear gradient formed from buffers A and B. Columns were run at room temperature at a flow rate of 1 ml/min.

E. Amino Acid Analysis

Amino acid analysis was carried out on a Beckman Model 6300 ion-exchange instrument following a 16 hr hydrolysis at 115° C. in 100 µl of 6 N HCl, 0.2% phenol that also contained 2 nM norleucine. Norleucine served as an internal standard to correct for losses that might occur during sample transfers and drying, etc. After hydrolysis, the HCl was dried

in a Speedvac and the resulting amino acids dissolved in 100 µl Beckman sample buffer that contained 2 nM homoserine that acted as a second internal standard to independently monitor transfer of the sample onto the analyzer. The instrument was calibrated with a 2 nM mixture of amino acids and was operated according to the manufacturer's programs and using the manufacturer's buffers. Data analysis was carried out on an external computer using Perkin Elmer/Nelson data acquisition software. Improved quantitation of cysteine was obtained by prior oxidation with performic acid in a second sample. This procedure converted both cysteine and cystine to cysteic acid. Performic acid oxidation may destroy tyrosine.

F. N-terminal Sequencing

Direct N-terminal amino acid sequencing of PX3.101 was carried out as previously described (Stone, et al, *In Techniques in Protein Chemistry*, Academic Press, 1992, New York, 23–34). PX3.101 purified by SDS-PAGE was electroblotted onto a Mini-ProbloTT pure PVDF membrane (Applied Biosystems, Foster City, Calif.). The band was visualized by Ponceau S and cut for N-terminal protein sequencing. Alternatively, PX3.101 from Sephadex G-50 chromatography was further purified and mutual derivatives were resolved by RP HPLC and HPLC-ion exchange chromatography as described above. N-terminal protein sequencing of either form was carried out by automated Edman degradation with an Applied Biosystems sequencer (Model Procise 494 cLc, Foster City, Calif.). An on-line HPLC-analyzer was used for the identification of phenylthiohydantoin (PTH) amino acids.

G. Internal Sequencing

Internal protein sequencing was performed according to standard methods (Stone, et al. *A Practical Guide to Protein and Peptide Purification for Microsequencing*, 2nd ed. Academic Press, 1993, New York, 43–69). A Coomassie Blue stained SDS-PAGE gel band was cut and subjected to in-gel trypsin digestion. Peptide fragments were subsequently extracted and analyzed by LC/MS/MS analysis. The resulted MS/MS spectra were compared with spectra in the National Center for Biotechnology Information (NCBI) non-redundant database to determine whether the protein was known. When no protein was identified using this method, the digest was purified by reverse phase HPLC. Detected peak fractions were collected and selected peptides were subjected to mass spectral analysis. N-terminal amino acid sequencing of these peptides was later carried out as described in Example I.

EXAMPLE II

Protein Purification and Characterization

A. Fractionation of Bee Venom

Lyophilized honeybee venom (approximately 0.5 g) (Aptironic Services, Richmond, British Columbia, Canada) or honeybee venom in suspension (approximate 0.5 g solid material per ml, Sigma, St. Louis, Mis.) was diluted in 10 ml of 0.1M ammonium formate buffer (pH 4.6) to give a solution having a concentration of approximately 0.05 g venom/ml. This solution was centrifuged and filtered through 0.45 µm filter, and then loaded onto a Sephadex G-50 column (two columns, each 1.5×170 cm (diameter×length) that were connected in series) pre-equilibrated with 0.1 M ammonium formate buffer (pH 4.6). The column was eluted at about 0.6 ml/min, and fractions of 100 drops (approximately 4.0 ml) were collected.

Fractions containing PX3.101 (fractions 65 to 72) appeared between the peaks of phospholipase A2 and melittin tetramer (FIG. 1). Fractions in the shoulder peak

(fractions 65 to 72) were analyzed by SDS gel electrophoresis. Only one major band was found to have molecular weight lower than phospholipase A2. Protein in this band had an apparent molecular weight of approximately 7700 daltons and was called PX3.101. Melittin monomer has a molecular weight of 3000 daltons.

B. Purification of PX3.101

Sephadex G-50 column fractions containing PX3.101 (as identified by SDS PAGE) were pooled and PX3.101 enriched by RP HPLC as described in Example I. A typical elution profile obtained during this RP HPLC step is shown in FIG. 2A. Fractions containing PX3.101 as confirmed by SDS-PAGE and trypsin inhibition assay (data not shown), were freeze-dried and further purified by ion-exchange HPLC as described in Example I. An elution profile is shown in FIG. 2B. A second RP HPLC served as the final step of the purification and was eluted according to the conditions described in Example I. As shown in the elution profile depicted in FIG. 2C, three major peaks were obtained, all of them with shoulders, in this final step. The main peaks were named Puri-#1, Puri-#2 and Puri-#3. These fractions had similar molecular weights based on SDS-PAGE, and all of them showed similar trypsin inhibition activity (data not shown). N-terminal sequencing of the main peaks confirmed that all of them were PX3.101 but with one to several N-terminal amino acids deleted (Table II). These molecules were considered mutual derivatives of PX3.101. These results are consistent with the direct N-terminal sequencing results in which the major PX3.101 band from the SDS-PAGE yielded multiple sequences (see below). The Puri-#1, Puri-#2 and Puri-#3 fractions were combined and used in the animal studies as well as mechanism of action studies of PX3.101 (see Examples V and VI).

C. Characterization

The major gel band from the SDS gel run with G-50 fractions was cut and subjected to in-gel trypsin digestion. The resulting digest was analyzed by LC/MS/MS. Ten to twelve major peptide fragments were formed as a result of digestion and their molecular weights determined (data not shown). No peptides in the NCBI non-redundant database were found to match the predicted molecular weights of peptides from trypsin digestion.

The peptide fragments from the trypsin digestion were then purified by RP-HPLC. Detailed protocols for the purification are described on the website (<http://info.med.vale.edu/wmkeck/prochem.htm>). The peak fractions (#47, 61, 62, 65, 75, 88 and 106) were collected and further analyzed by mass spectroscopy. Four peptide fragments (fractions #47, 62, 75 and 88) were selected for amino acid sequencing. The amino acid sequences of the peptides are shown in Table I below.

Direct N-terminal sequencing of the major SDS gel band blotted to PVDF membrane was also tried, but yielded multiple amino acids in each cycle of the sequencing (data not shown). This result suggested either major contaminants or the presence of multiple forms of PX3.101 that were mutual derivatives. Further study indicated the presence of multiple forms of this molecule resulted from deletions of different numbers of amino acids from the N-terminus (see below).

Polypeptides contained in collected Fractions from the final RP-HPLC column (i.e., Puri-#1, Puri-#2, and Puri-#3; see FIG. 2C), had similar molecular weights based on SDS-PAGE, and all of them showed similar trypsin inhibition activity (data not shown). N-terminal sequencing of the main peaks confirmed that all of them were PX3.101 but with one to several N-terminal amino acids deleted (Table

II). These molecules were considered mutual derivatives of PX3.101. These results are consistent with the direct N-terminal sequencing results in which the major PX3.101 band from the SDS-PAGE yielded multiple sequences (see above).

A predicted molecular weight of about 7,700 daltons correlated well with the migration of PX3.101 on SDS-gel but did not correspond well with where the protein eluted in the elution profile for the Sephadex G-50 sizing column. The fact that the fraction containing PX3.101 eluted between Phospholipase A2 (MW 19,000) and melittin tetramer (11,400) indicated an apparent molecular weight between 11,400 and 19,000. This discrepancy suggested that either the PX3.101 was present as a dimer, or that there was significant post-translational modification of the protein. Gel electrophoresis results obtained under non-reducing condition indicated the presence of dimers of PX3.101 molecules (data not shown).

A glucose assay did not show any glycosylation of PX3.101 (data not shown). The molecular weights of Puri-#1, Puri-#2 and Puri-#3, as determined by mass spectral analysis matched well with the molecular weights predicted from their amino acid sequences (Table II). These results suggested that there are no post-translational modifications of these derivatives. The C-terminals of these mutual derivatives are likely to be intact and to be the same.

The amino acid analysis of the purified PX3.101 used for animal studies showed that it had an extinction coefficient at 280 nm of 0.286 ml/mg/cm. Protein amount in any given preparation was determined by its absorbance at 280 nm and the extinction coefficient.

EXAMPLE III

Cloning PX3.101 cDNA and Predicted Protein Sequence

A. Method

The degenerate oligonucleotide primer (5' ATGGATC-CAAYGARATHTTYWSNAG 3'—SEQ ID NO:8) Y=C or T; R=A or G; H=A or C or T; W=A or T; N=A or C or G or T) was designed based on the amino acid sequence (NEIFSR—SEQ ID NO:9) obtained from protein sequencing. All the PX3.101 primers and Oligo(dT) (5'TTGCGGCCGCTTTTITTTTITTTTITTTT3'—SEQ ID NO:10) were synthesized and purified by Genemed Synthesis, Inc.

The total RNA of the honeybee venom gland was prepared as previously described (Chomczynski, et al, (1995) *Anal Biochem.* 225:163). Venom glands were collected from honeybee *Apis mellifera* by Apitronic Services (Richmond, British Columbia, Canada) and stored at -80° C. 100 venom glands in STAT-60 solution (TEL-TEST "B" Inc. TX) were homogenized using a glass-Teflon homogenizer. Total RNA was extracted and isolated using the RNA isolation kit from TEL-TEST "B" Inc.

The first strand cDNA was synthesized by reverse transcription using the total RNA from honey bee gland as the template and Oligo(dT) as the primer. The PX3.101 gene fragment was amplified by PCR using the degenerate PX3.101 primer and oligo(dT). The amplified DNA fragments were cloned into the NotI and BamHI sites of pBlue-script sk(+) vector. The sequences of the DNA fragments were obtained through contracted service from Genemed Syntheses, Inc. The predicted protein sequences were analyzed to see if they matched peptide sequences obtained through protein sequencing: i.e., sequences: a) PSNEIFSR (SEQ ID NO:11) (residues 38 to 45 of SEQ ID NO:2), b) GFGGFGGLGGR (SEQ ID NO:12) (residues 24 to 34 of

SEQ ID NO:2), c) VCVPR (SEQ ID NO:13) (residues 84 to 88 of SEQ ID NO:2), or d) PNVVPK (SEQ ID NO:14) (residues 55–60 of SEQ ID NO:2).

To get the full-length cDNA of PX3.101 gene including the coding sequence for the amino terminals and the signal peptide, 5'-RACE (Rapid Amplification of cDNA End) system (Gibco, MD) was used. Oligonucleotide primers ASEQ2 (5' ATCGCGGAACGCA 3'—SEQ ID NO:15) and ASEQ3 (5' AAGGATCCAAGTCTACATACAC 3'—SEQ ID NO:16) were synthesized and used to amplified the 5' end of PX3.101 cDNA. The PCR products were cloned into BamHI and Sall sites of pBluescript sk(+) vector and sequenced. The predicted protein sequences were analyzed to see if they matched the peptide sequence obtained through protein sequencing: GFGGFGGLGGR (SEQ ID NO:12) (residues 24 to 34 of SEQ ID NO:2). The protein and gene sequences of PX3.101 were analyzed by searching the database to identify any structural or functional motifs.

B. Results

A DNA fragment, containing coding sequence for peptides (NEIFSR (SEQ ID NO:10)—residues 40 to 45 of SEQ ID NO:2), (VCVPR (SEQ ID NO:13)—residues 84 to 88 of SEQ ID NO:2), and (PNVVPK (SEQ ID NO:14)—residues 55 to 60 of SEQ ID NO:2), was discovered. This DNA fragment is part of PX3.101 gene. The full-length PX3.101 gene (472 base pairs; SEQ ID NO:1) was isolated from the honeybee cDNA library. It contains a 276 base pair coding sequence (residues 74 to 349 of SEQ ID NO:1, a 5' end untranslated region, a 3' end untranslated region, and a poly(A) tail. The predicted PX3.101 protein consists of 92 amino acids (SEQ ID NO:2), including peptide (GFGGFGGLGGR (SEQ ID NO:12)—residues 24 to 34 of SEQ ID NO:2). The nucleotide and predicted amino acid sequence of PX3.101 are shown in FIG. 3A.

Like other secreted molecules, PX3.101 protein consists of a 19 amino acid signal peptide at the N-terminus (FIG. 3B; residues 1–19 of SEQ ID NO:2). The coding sequence for the PX3.101 signal peptide can be used to construct expression vector, to express recombinant proteins in secreted form.

The secreted and natural PX3.101 protein in honeybee venom starts with five GGX repeats (FIG. 3B; residues 20 to 34 of SEQ ID NO:2). GGX repeats are present in several structure proteins, including keratin (CAA28991), abducin (2739489), fibrillarin (P22232), elastin (207462), spider silk protein (AAC38847), procollagen D (2772914) and procollagen P (2388676) of mussel byssus, homeotic protein Spalt-accessory (AAC38847), putative immediate early protein of Alcelaphine herpesvirus 1 (2338034), EBNA-1 of Epstein-Barr virus (P0321 1), and many Mycobacterium tuberculosis proteins (CAA17751, CAA15537, CAA17576, CAA17749). GGX repeats form a condensed helical structure and may be involved in formation of polymers. Interestingly, auto-antibodies against keratin, fibrillarin or elastin are found in rheumatoid arthritis patients.

The C-terminus of PX3.101 is cysteine rich (FIG. 3B; residues 35–92 of SEQ ID NO:2). In 58 amino acids, there are 10 cysteines. A cysteine-rich motif such as this is present in a group of proteins, including tectorin (CAA68138), zonadhesin (3327421), IgG Fc binding protein (AAD15624), von Willebrand factor (CAA27765), ECM 18 (1100979), mucin (AF015521), hemocytin (P980920), SCO-spodin (CAA69868), tumor necrosis factor receptor II (2739045), a chymotrypsin inhibitor from honeybee (4699856), anti-coagulant protein C2 (1203803), and several proteins with similarity to EGF-like domain (CAA98455, AF016450, 1226303, U70857, 1226304).

Most of the proteins above are extracellular proteins mediating different signal transduction pathways and have more than one cysteine-rich motif. Tectorin, the protein associated with hearing disability, has three such cysteine-rich domains. IgG Fc binding protein has as many as twelve.

A database search identified several protein candidates as potential homologues of PX3.101. All of them are small proteins with a signal peptide and at least one cysteine-rich motif. Their schematic structures and name or accession number are included in FIG. 3C.

EXAMPLE IV

Expression and Purification of Recombinant PX3.101 Protein

A. Method

To generate the recombinant PX3.101 protein, the full-length cDNA of PX3.101 was cloned into pFastBachTb, a baculovirus expression vector, in-frame with coding sequence for His-tag (Gibco, MD) (see FIG. 8). This virus expression vector was designed for high-level productions and rapid purification of the recombinant protein (Lukow, et al, Bio/Technology, 1989, 6:47).

pFastBachTb-PX3.101 was used to transform DH10Bac cells (Gibco) for transposition into the bacmid. The recombinant bacid containing PX3.101 cDNA was isolated and assessed by PCR. To generate the baculovirus, SF9 cells (Invitrogen, Calif.) were infected with the recombinant bacmid. Following 5 days of incubation at 30° C., the virus stock was collected and used to infect SF9 cells to generate high titer virus stock. The titer of virus stock was determined by plaque assay.

To optimize the condition to generate PX3.101 protein, the recombinant baculovirus stock was used to infect High Five cells (Invitrogen, Calif.) an insect cell line generally expressing significantly higher levels of recombinant proteins compared to other insect cells (Wickham, et al, *Bio-technol. Prog.*, 1992, 8:391). High-five cells in mid-log phase of growth in one liter of serum-free medium were infected with the recombinant baculovirus stock (1:100 v/v). After incubation at 30° C. for 96 hrs, cells were harvested by centrifugation and lysed using guanidinium lysis buffer (6M guanidine hydrochloride, 20 mM sodium phosphate, 500 mM sodium chloride, pH 7.8).

After centrifugation, the supernatant was collected and incubated with pre-equilibrated PROBOND resin (Invitrogen, Calif.) at 4° C. for 3 hours. The column was washed sequentially with two bed volumes of the following buffers twice: denaturing binding buffer (8 M urea, 20 mM sodium phosphate, 500 mM sodium chloride, pH 7.8), denaturing wash buffer 1 (8 M urea, 20 mM sodium phosphate, 500 mM sodium chloride, pH 6.0), and denaturing wash buffer 2 (8 M urea, 20 mM sodium phosphate, 500 mM sodium chloride, pH 5.3). Recombinant PX3.101 was eluted from the column using denaturing elution buffer (8 M urea, 20 mM sodium phosphate, 500 mM sodium chloride, pH 4.0). The His tag was removed by rTEV protease digestion.

B. Results

Recombinant PX3.101 is expressed and soluble in denaturing lysis buffer. After removal of His tag by protease rTEV digestion, recombinant PX3.101 is almost identical in size on SDS-polyacrylamide gel electrophoresis. Recombinant PX3.101 was purified using PROBOND affinity column followed by HPLC. About 20 mg PX3.101 protein was obtained from 1 liter of cell culture.

After protein refolding, recombinant PX3.101 protein (●) and natural PX3.101 protein (▲) show equivalent activities in inhibiting the binding of IL-8 to its receptor CXCR2 (FIG. 7B).

EXAMPLE V

Animal Studies—Collagen Induced Arthritis Mouse Studies
A. General Method

The collagen-Induced Arthritis (CIA) animal model is widely acknowledged as the most appropriate in vivo model system to test potential therapeutics to treat rheumatoid arthritis and is recommended by the Food and Drug Administration for pre-clinical testing in preparation for an IND (Investigational New Drug) filing.

The following two animal studies began with 8–10 week old DBA/1J male mice from Jackson Laboratories (Bar Harbor, Me.). Disease was induced in all animals following the protocol described by Rosloniec, et al. (*Current Protocols in Immunology*, John Wiley & Sons, Inc. 1996). In brief, chicken collagen type II was dissolved in 10 mM acetic acid at 4 mg/ml and stirred overnight at 4° C. It is important that native collagen type II be kept cold while being dissolved to prevent its denaturation. Using a high-speed homogenizer, chicken type II collagen was emulsified in an equal volume of Complete Freund's Adjuvant (CFA) just prior to immunization. The solution is kept cold throughout the emulsification.

On Day 1, DBA/1JLacJ mice were injected interdermally at the base of the tail with 0.1 mg of the chicken collagen Type II emulsified in CFA. On Day 21, a second identical injection was administered.

As set forth below, various treatments were administered subcutaneously, at various times into tissue of the upper back/shoulder area. Inflammation was recorded throughout each study, at least twice weekly, by counting the number of swollen toes, paws and ankles of each animal. Each joint was assigned a score of either 0 (no inflammation) or 1 (inflammation). According to this scoring system, a maximum score of 28 (7 measurements per limb×4 limbs) and a minimum score of 0 could be assigned to an animal at any single scoring occasion. This number, is representative of the disease Severity.

B. Study 1

1. Method

Five groups of 10 mice each were treated as follows:

Group 1: PX3.101 (200 µg/kg) administered subcutaneously for 15 days starting on Day 6.

Group 2: Bee Venom (1000 µg/kg) (obtained from Apitronic Services and dissolved in PBS administered subcutaneously for 30 days starting on Day 1.

Group 3: Negative Control (Phosphate Buffered in normal Saline (PBS) administered subcutaneously for 30 days starting on Day 1).

Group 4: INDOMETHACIN (positive control; available from Sigma Chemical Co., St. Louis, Mo.) administered orally for 30 days starting on Day 1 (1000 µg/kg).

Group 5: Normal Control (same as Group #3, except mice received no collagen).

On Day 52, blood samples were obtained from the Negative Control (Group #3) and PX3.101 treated animals (Group #1) to evaluate the immunogenicity of PX3.101 in this animal model. Serum was prepared and tested by Western Blot analysis as described in Example I.

2. Results

Daily treatment of mice with PX3.101 at the does of 200 µg/kg from Day 16 to Day 30 (Group #3) suppressed inflammation in CIA (collagen-induced arthritis) mice. Its activity was comparable to INDOMETHACIN, a known anti-arthritis drug (daily treatment at a does of 1 mg/kg, Group #4) (see FIG. 4A). Statistical significance when compared to the Negative Control Group #3 is $p < 0.05$.

Histopathologic studies of joints from the mice treated with PX3.101 and PBS further demonstrated the therapeutic activity of PX3.101 in suppressing inflammation (see FIGS. 5A–5C). In FIGS. 5A–5C, the white space located near the center of the photographs is a space between the bones (large dark regions) in a joint of the mouse that contains synovial fluid. The small dark spots or granules, particularly noticeable in FIG. 5B, are the nuclei of leukocytes (e.g., neutrophils, T-cells, macrophages, and other cells stimulated as part of an immune response) that have infiltrated the joint. These leukocytes actively degrade bone.

FIG. 5A shows a normal joint wherein collagen has not been injected to induce arthritis (Group 5). The dark bony material is smooth and undegraded and there are very few leukocytes present. In sharp contrast, many leukocytes were present in the joint from mice which were injected with collagen and then treated with PBS (Group 3, see FIG. 5B). In this negative control treatment group, bone erosion and penetration by leukocytes was observed and cartilage damage was obvious (see FIG. 5B). FIG. 5C is a photograph of a joint from a mouse from Group 1 that was treated with PX3.101. Very little bone erosion and cartilage damage was observed. There are also very few leukocytes present in the joint. This result suggests that PX3.101 can inhibit migration of leukocytes to inflammatory sites. The hypothesis is supported by our findings that PX3.101 inhibits the interaction between chemokine IL-8 and its receptor CXCR1 and CXCR2 (see FIG. 7A and Example VI). IL-8 is the major chemokine involved in inflammation. Its function includes recruiting neutrophils to the inflammatory site and activating them to release superoxide, proteases, and bioactive lipids.

Western Blot analysis did not detect antibodies against PX3.101 in the serum of mice treated with PX3.101 at 200 µg/kg for 15 consecutive days.

C. Study 2

1. Method

Four groups of 7–8 mice each were treated as follows:

Group 1: PX3.101 (200 µg/kg) administered subcutaneously for 15 days starting on Day 22.

Group 2: PX3.101 (40 µg/kg) administered subcutaneously for 15 days starting on Day 22.

Group 3: PX3.101 (8 µg/kg) administered subcutaneously for 15 days starting on Day 22.

Group 4: Negative Control (Phosphate Buffered in normal Saline (PBS) administered subcutaneously for 15 days starting on Day 22).

2. Results

The effectiveness of various concentrations of PX3.101 in suppressing inflammation in CIA (collagen-induced arthritis) was demonstrated. In this study, mice received PX3.101 treatment from Day 22 to Day 37, instead of Day 16 to Day 30 as in Study 1 above. Among three dosages (8 µg/kg, 40 µg/kg, and 200 µg/kg) tested, 40 µg/kg appears to be the optimal concentration (FIG. 6).

Results from this study indicate that the activities of PX3.101 molecule in this animal model depend on its dosage in a non-linear manner. Such phenomena have been observed in many cases in pre-clinical or clinical investigations, for example TNF-α soluble receptor and relaxin, a drug candidate in late stage clinical development for Scleroderma.

In addition, a substantial sustained therapeutic effect was observed after the treatment with PX3.101 was discontinued, suggesting possible long-lasting effect of this molecule.

D. Summary of Animal Studies

Therapeutic potential was demonstrated for a purified component from honeybee venom identified as PX3.101. In

evaluating the data from the two studies, it appears that the in vivo activity of PX3.101 is dependent on its dosage and the time to start treatment. The most effective dosage of PX3.101 of the dosages tested was 40 $\mu\text{g/kg}$.

EXAMPLE VI

Mechanism of Action Studies

A. Method—Chemokine or Cytokine/Receptor Binding

Experiments to examine the effects of PX3.101 on chemokine or cytokine/receptor binding were carried out by Panlabs. For the inhibition of IL-8/CXCR2 binding, purified (naturally-occurring or recombinant) PX3.101 was added to 0.2 ml reaction solution that contained 0.15 mg/ml of a membrane preparation of human recombinant CHO cells expressing CXCR2, 0.015 nM ^{125}I -labeled IL-8, and 10 nM unlabeled IL-8 to give a final PX3.101 concentration of 0, 0.01, 0.1, 1.0 and 10 μM . Reaction mixtures were incubated for 60 minutes at room temperature. Bound radioligand was then separated from unbound radioligand and the radioactivity measured on a gamma counter. Similar experiments were carried out to examine the effect of PX3.101 on IL-8/CXCR1 interaction, where the membrane preparation of human recombinant CHO cells that expressed CXCR1 was used and a single dose of PX3.101 (10 μM) was tested. TNF- α /TNF- α receptor binding experiments were carried out in a similar manner. Briefly, 10 μM PX3.101 was added to the reaction solution that contained 50 mM Tris-HCl (pH 7.4), 0.5 mM EDTA, 0.028 nM ^{125}I -labeled TNF- α , 40 nM unlabeled TNF- α , and preparation of human U937 cells that expressed TNF- α receptor. The mixture was allowed to incubate for 3 hours at 4° C. Bound radioligands were separated from unbound and the radioactivity was counted on the gamma counter.

B. Results

PX3.101 was found to specifically inhibit the interaction between IL-8 and CXCR1 and the interaction between IL-8 and CXCR2 (FIG. 7A). PX3.101 inhibited IL-8 and CXCR2 interaction in a dose-dependent manner (FIG. 7A). Preliminary tests show an IC₅₀ of 0.5 μM . However, the binding of TNF- α to its receptor was not affected by PX3.101 (FIG. 7A).

IL-8 is a major chemokine that regulates the inflammatory process. There is also research suggesting it may also be involved in tumor angiogenesis and tumor metastasis (Koch, et al., (1992) *Science* 258:1798). Since PX3.101 inhibits the binding of IL-8 to its receptors CXCR1 and CXCR2, PX3.101 is expected to be effective in the treatment of cancer, inflammatory diseases, autoimmune diseases and other diseases involving IL-8. Inhibition of the IL-8/CXCR2 interaction by purified naturally-occurring PX3.101 and recombinant PX3.101 is shown in FIG. 7B.

PX3.101 was also found to inhibit several enzymes involved in the pathogenesis of rheumatoid arthritis, including cyclooxygenases (COX 1 and COX2), phospholipase A2, phospholipase C, lipoxygenase, and the proteases

trypsin and cathepsin G (data not shown). Several of these enzymes are either integrated in the phospholipid membrane (cyclooxygenases) or use fatty acids or phospholipids as their substrates (phospholipase A2, phospholipase C, lipoxygenase). Interestingly, PX3.101 inhibited COX1 when the enzyme purified in lipid vesicles was used but showed no inhibition to the free enzyme in solution. This result suggests that the inhibitions of the lipid/fatty acid related enzymes might occur through non-specific interaction between PX3.101 and lipids/fatty acids.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

TABLES

TABLE I

Selected fragments resulting from in-gel trypsin digestion			
Peptide Fragment	SEQ ID NO:	Amino Acid Sequence	MW (D) by Mass Spec
#47	17	V-X-V-P-R	Not Determined
#62	18	X-P-S-N-E-I-F-S-R	1124.2
#75	12	G-F-G-G-F-G-G-L-G-G-R	982.9
#88	19	X-X-P-N-V-V-P-K	Not Determined

*V = Valine, P = Proline, C = Cysteine, K = Lysine, S = Serine, I = Isoleucine, E = Glutamic acid, N = Asparagine, X = Unsure

TABLE II

N-Terminal Sequences for PX3.101 protein fractions from honey bee venom (FIG. 2C)				
Name	SEQ ID NO:	N-terminal Sequence of PX3.101 proteins in honey bee venom	Measured MW (D) by mass spec	Predicted MW (D)* based on sequence
Puri-#1	20	G-G-F-G-G-L-G-G-R-G	7178	7178
Puri-#2	21	G-F-G-G-F-G-G-L-G-G	7405	7382
Puri-#3	22	F-G-G-F-G-G-F-G-G-L	7586	7586

*The molecular weights were predicted using IntelliGenetics program assuming that all the 10 cysteines form 5 pairs of disulfide bonds. It is also assumed that the C-terminus of the protein is intact and that there are no post-translational modifications of the molecule.
G = Glycine, F = Phenylalanine, R = Arginine, L = Leucine,

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 22

<210> SEQ ID NO 1

<211> LENGTH: 472

<212> TYPE: DNA

-continued

```

<213> ORGANISM: Apis mellifera
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (74)..(352)
<223> OTHER INFORMATION: honey bee venom PK3.101 protein

<400> SEQUENCE: 1

attcacagtg caacgtaagt tctttcttc ttttttttt cgaaaaaaca actttgtttg      60
agaagaacaa aac atg tct cgt ctg gtt ctt gcc tcc ttc ctt ctt ttg      109
      Met Ser Arg Leu Val Leu Ala Ser Phe Leu Leu Leu
              1              5              10

gca att ttc tcc atg ctt gtt gga gga ttt gga gga ttt gga gga ttt      157
Ala Ile Phe Ser Met Leu Val Gly Gly Phe Gly Gly Phe Gly Gly Phe
      15              20              25

gga gga ctt gga gga cgt ggt aaa tgt cca agc aat gag atc ttc agt      205
Gly Gly Leu Gly Gly Arg Gly Lys Cys Pro Ser Asn Glu Ile Phe Ser
      30              35              40

aga tgc gat gga cgg tgc caa cgt ttt tgc ccc aat gtt gtt cct aaa      253
Arg Cys Asp Gly Arg Cys Gln Arg Phe Cys Pro Asn Val Val Pro Lys
      45              50              55              60

cct tta tgc atc aag ata tgt gca cca gga tgt gta tgt aga ctt ggt      301
Pro Leu Cys Ile Lys Ile Cys Ala Pro Gly Cys Val Cys Arg Leu Gly
      65              70              75

tat tta agg aat aaa aag aag gta tgc gtt ccg cga tct aaa tgc gga      349
Tyr Leu Arg Asn Lys Lys Lys Val Cys Val Pro Arg Ser Lys Cys Gly
      80              85              90

tgacttttat aattattttca tgattatttt atgattgttt aacaattatt gtattgtatt      409
ttatcattca taaaaattgt tatgttatta ttttatcagt aaaaaaaaaa aaaaaaaaaa      469
aaa                                                                 472

```

```

<210> SEQ ID NO 2
<211> LENGTH: 92
<212> TYPE: PRT
<213> ORGANISM: Apis mellifera

```

```

<400> SEQUENCE: 2

Met Ser Arg Leu Val Leu Ala Ser Phe Leu Leu Leu Ala Ile Phe Ser
      1              5              10              15

Met Leu Val Gly Gly Phe Gly Gly Phe Gly Gly Phe Gly Gly Leu Gly
      20              25              30

Gly Arg Gly Lys Cys Pro Ser Asn Glu Ile Phe Ser Arg Cys Asp Gly
      35              40              45

Arg Cys Gln Arg Phe Cys Pro Asn Val Val Pro Lys Pro Leu Cys Ile
      50              55              60

Lys Ile Cys Ala Pro Gly Cys Val Cys Arg Leu Gly Tyr Leu Arg Asn
      65              70              75              80

Lys Lys Lys Val Cys Val Pro Arg Ser Lys Cys Gly
      85              90

```

```

<210> SEQ ID NO 3
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:forward
      primer ASEQ10

```

```

<400> SEQUENCE: 3

aaggatccac agtgcaacgt aagttc      26

```

-continued

<210> SEQ ID NO 4
<211> LENGTH: 55
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:forward
primer ASEQ11

<400> SEQUENCE: 4

aaggatccgg aggatttggg ggatttggag gatttggagg acttggagga cgtgg 55

<210> SEQ ID NO 5
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:reverse
primer ASEQ13

<400> SEQUENCE: 5

actgataaaa taataac 17

<210> SEQ ID NO 6
<211> LENGTH: 16
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:reverse
primer ASEQ14

<400> SEQUENCE: 6

atgaatgata aaatac 16

<210> SEQ ID NO 7
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:reverse
primer ASEQ15

<400> SEQUENCE: 7

ttataaaagt catcgc 17

<210> SEQ ID NO 8
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:degenerate
oligonucleotide primer
<221> NAME/KEY: modified_base
<222> LOCATION: (23)
<223> OTHER INFORMATION: n = g, a, c or t

<400> SEQUENCE: 8

atggatccaa ygarathtty wsnag 25

<210> SEQ ID NO 9
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:amino acid
sequence obtained from protein sequencing

<400> SEQUENCE: 9

-continued

Asn Glu Ile Phe Ser Arg
1 5

<210> SEQ ID NO 10
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:Oligo(dT)

<400> SEQUENCE: 10

ttgcggccgc tttttttttt tttttttt

28

<210> SEQ ID NO 11
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:residues
38-45 of SEQ ID NO:2 obtained through protein
sequencing; peptide fragment #75 from in-gel
trypsin digestion

<400> SEQUENCE: 11

Pro Ser Asn Glu Ile Phe Ser Arg
1 5

<210> SEQ ID NO 12
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:residues
24-34 of SEQ ID NO:2 obtained through protein sequencing

<400> SEQUENCE: 12

Gly Phe Gly Gly Phe Gly Gly Leu Gly Gly Arg
1 5 10

<210> SEQ ID NO 13
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:residues
84-88 of SEQ ID NO:2 obtained through protein sequencing

<400> SEQUENCE: 13

Val Cys Val Pro Arg
1 5

<210> SEQ ID NO 14
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:residues
55-60 of SEQ ID NO:2 obtained through protein sequencing

<400> SEQUENCE: 14

Pro Asn Val Val Pro Lys
1 5

<210> SEQ ID NO 15
<211> LENGTH: 13
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: Description of Artificial Sequence:5' RACE
oligonucleotide primer ASEQ2

<400> SEQUENCE: 15

atcgcggaac gca 13

<210> SEQ ID NO 16
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:5' RACE
oligonucleotide primer ASEQ3

<400> SEQUENCE: 16

aaggatccaa gtctacatac ac 22

<210> SEQ ID NO 17
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:peptide
fragment #47 from in-gel trypsin digestion
<221> NAME/KEY: MOD_RES
<222> LOCATION: (2)
<223> OTHER INFORMATION: Xaa = unsure amino acid

<400> SEQUENCE: 17

Val Xaa Val Pro Arg
1 5

<210> SEQ ID NO 18
<211> LENGTH: 9
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:peptide
fragment #62 from in-gel trypsin digestion
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)
<223> OTHER INFORMATION: Xaa = unsure amino acid

<400> SEQUENCE: 18

Xaa Pro Ser Asn Glu Ile Phe Ser Arg
1 5

<210> SEQ ID NO 19
<211> LENGTH: 8
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:peptide
fragment #88 from in-gel trypsin digestion
<221> NAME/KEY: MOD_RES
<222> LOCATION: (1)..(2)
<223> OTHER INFORMATION: Xaa = unsure amino acid

<400> SEQUENCE: 19

Xaa Xaa Pro Asn Val Val Pro Lys
1 5

<210> SEQ ID NO 20
<211> LENGTH: 10
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence:N-terminal
sequence from Puri-#1

-continued

<400> SEQUENCE: 20

Gly Gly Phe Gly Gly Leu Gly Gly Arg Gly
 1 5 10

<210> SEQ ID NO 21

<211> LENGTH: 10

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence:N-terminal sequence from Puri-#2

<400> SEQUENCE: 21

Gly Phe Gly Gly Phe Gly Gly Leu Gly Gly
 1 5 10

<210> SEQ ID NO 22

<211> LENGTH: 10

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence:N-terminal sequence from Puri-#3

<400> SEQUENCE: 22

Phe Gly Gly Phe Gly Gly Phe Gly Gly Leu
 1 5 10

What is claimed is:

1. An isolated nucleic acid molecule that comprises a polynucleotide sequence that encodes a polypeptide having an amino acid sequence at least 75% identical to an amino acid sequence as set forth in SEQ ID NO:2 over a region at least about 40 amino acids in length when compared using the BLASTP algorithm with a wordlength (W) of 3, and the BLOSUM62 scoring matrix, wherein the polypeptide is effective to reduce the symptoms of an inflammatory disease.
2. The nucleic acid of claim 1, wherein the polynucleotide sequence encodes a polypeptide having an amino acid sequence as shown in SEQ ID NO:2.
3. An isolated nucleic acid molecule that comprises a polynucleotide sequence at least 75% identical to a nucleic acid sequence set forth in nucleotides 74 to 349 of SEQ ID NO:1 over a region of at least 50 nucleotides in length when compared using the BLASTN algorithm with a wordlength (W) of 11, M=5, and N=-4 and encodes a polypeptide that is effective in reducing the symptoms of an inflammatory disease.
4. The nucleic acid of claim 3, wherein the inflammatory disease is rheumatoid arthritis.
5. The nucleic acid of claim 1, wherein the polynucleotide sequence hybridizes to a nucleic acid having a sequence as set forth in residues 74 to 349 of SEQ ID NO:1 under stringent conditions, wherein the stringent conditions are conditions in which the ionic strength is equivalent to a solution containing 0.01 to 0.1M sodium ion, the pH is pH 7.0 to 8.3 and the temperature is at least 30° C. for polynucleotides 10 to 50 nucleotides in length and at least 60° C. for polynucleotides greater than 50 nucleotides in length.
6. The nucleic acid of claim 1, wherein the polynucleotide sequence is as set forth in residues 74 to 349 of SEQ ID NO:1.
7. The nucleic acid of claim 1, wherein the polynucleotide sequence is one that can be amplified using the forward primer 5' AAGGATCCACAGTGCAACGTAAGTTC 3' (SEQ ID NO:3) and reverse primer 5' ACT-GATAAAATAATAAC 3' (SEQ ID NO:5).
8. The nucleic acid of claim 1, wherein the polynucleotide sequence is as set forth in SEQ ID NO:1.
9. The nucleic acid of claim 1, wherein the polynucleotide sequence is derived from a sample from bee venom.
10. The nucleic acid of claim 1, further comprising a promoter sequence operably linked to the polynucleotide sequence.
11. A vector comprising a nucleic acid of claim 1.
12. The vector of claim 11, wherein said vector is a baculovirus.
13. A cell containing the vector of claim 11.
14. A cell comprising a recombinant expression cassette comprising a promoter operably linked to a polynucleotide sequence which is at least about 75% identical to residues 74 to 349 of the polynucleotide sequence as set forth in SEQ ID NO:1 over a region at least about 50 nucleotides in length when compared using the BLASTN algorithm with a wordlength (W) of 11, M=5, and N=-4 and which encodes a polypeptide that is effective in reducing the symptoms of an inflammatory disease.
15. The cell of claim 14, wherein the insect cell line is a High Five insect cell line.
16. The nucleic acid of claim 14, wherein the inflammatory disease is rheumatoid arthritis.
17. The cell of claim 13, wherein the polynucleotide hybridizes to a nucleic acid having a sequence as set forth in residues 74 to 349 of SEQ ID NO:1 under stringent conditions, wherein the stringent conditions are conditions in which the ionic strength is equivalent to a solution containing 0.01 to 0.1M sodium ion, the pH is pH 7.0 to 8.3 and the temperature is at least 30° C. for polynucleotides 10

47

to 50 nucleotides in length and at least 60° C. for polynucleotides greater than 50 nucleotides in length.

18. The cell of claim 13, wherein the polynucleotide sequence is as set forth in residues 74 to 349 of SEQ ID NO:1.

19. The cell of claim 13, wherein the cell is an insect cell line.

20. A method for producing a polypeptide, the method comprising the steps of:

(a) culturing a host cell containing the nucleic acid of claim 1 under conditions suitable for the expression of the polypeptide; and

(b) recovering the polypeptide from the host cell culture.

21. The nucleic acid of claim 1, wherein the inflammatory disease is rheumatoid arthritis.

22. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of

48

(a) a deoxyribonucleotide sequence complementary to nucleotides 74 to 349 of SEQ ID NO:1;

(b) a ribonucleotide sequence complementary to nucleotides 74 to 349 of SEQ ID NO:1;

(c) a nucleotide sequence complementary to the deoxyribonucleotide sequence of (a) or to the ribonucleotide sequence of (b);

(d) a nucleotide sequence of at least 23 consecutive nucleotides that hybridizes to nucleotides 74 to 349 of SEQ ID NO:1; and

(e) a nucleotide sequence that hybridizes to a nucleotide sequence of (d).

23. A vector comprising the nucleic acid of claim 21.

24. A cell containing the vector of claim 23.

* * * * *

OTHER PUBLICATIONS

Eiseman, J. L.; von Bredow, J. and Alvares, A.P.; Effect Of Honeybee (*Apis Mellifera*) Venom On The Course Of Adjuvant-Induced Arthritis And Depression Of Drug Metabolism In The Rat, *Biochemical Pharmacology*, 1982, vol. 31, No. 6, pp. 1139-1146.

Beck, Bodog F., M. D.; Bee Venom Therapy (Bee Venom, Its Nature, and Its Effect on Arthritic and Rheumatoid Conditions), D. Appleton-Century Company, Inc., New York and London, 1935, pp. 150-197.

Broadman, Joseph, M.D., Bee Venom (The Natural Curative For Arthritis And Rheumatism), G. P. Putnam's Sons, New York, 1962, pp. 13-18.

Broadman, Joseph, M.D., Bee Venom (The Natural Curative For Arthritis And Rheumatism), G. P. Putnam's Sons, New York, 1962, pp. 28-50.

McCulloch, Michael, L.Ac., Frequently Asked Questions about Apitherapy, <http://www.apitherapy.org/aas/faq.html>, 1997, pp. 1-9.

Frei, Erich; Schuh, Reinhard; Baumgartner, Stefan; Burri, Maya; Noll, Markus; Jurgens, Gerd; Seifert, Eveline; Nauber, Ulrich and Jackie, Herbert; Molecular characterization of spalt, a homoeotic gene required for head and tail development in the *Drosophila* embryo, *The EMBO Journal*, 1988, vol. 7, No. 1, pp. 197-204.



US006025156A

United States Patent [19]**Gwynn et al.**[11] **Patent Number:** **6,025,156**[45] **Date of Patent:** **Feb. 15, 2000**[54] **TOPOISOMERASE III**

[75] Inventors: **Michael N. Gwynn**, Chester Springs;
Howard Kallendar, King of Prussia;
Leslie M. Palmer, Malvern, all of Pa.

[73] Assignee: **Smithkline Beecham Corporation**,
 Philadelphia, Pa.

[21] Appl. No.: **08/949,588**

[22] Filed: **Oct. 14, 1997**

Related U.S. Application Data

[60] Provisional application No. 60/028,417, Oct. 15, 1996.

[51] Int. Cl.⁷ **C07H 21/04**

[52] U.S. Cl. **435/69.1; 536/23.7; 536/23.2;**
536/24.32; 435/183; 435/320.1; 435/252.3;
435/471; 435/233; 530/350

[58] Field of Search **536/23.2, 23.7,**
536/24.32; 435/69.1, 252.3, 320.1, 233,
471, 183; 530/350

[56] **References Cited****U.S. PATENT DOCUMENTS**

5,004,758 4/1991 Boehm et al. 514/283

OTHER PUBLICATIONS

Lerner, et al.: "High Throughput Screen for Inhibitors of Bacterial DNA Topoisomerase I Using the Scintillation Proximity Assay" *Journal of Biomolecular Screening*, vol. 1, No. 3; pp. 135-143 (1996).

Snewin, et al.: "Cloning and Characterisation of a gene from *Plasmodium vivax* and *P. knowlesi*: homology with valine-tRNA synthetase" *Gene* vol. 173 No. 2 (1996) pp. 137-145.

Brown, et al.: "A Sign Inversion Mechanism for Enzymatic Supercoiling of DNA" *Science* vol. 206, pp. 1081-1083 (1979).

Yoshimura, et al.: "*Bacillus subtilis* secretes a foreign protein by the signal sequence of *Bacillus amyloliquefaciens* neutral protease" *Appl. Microbiol.* vol. 23, pp. 250-256 (1986).

Huang, et al.: "DNA Topology and Its Biological Effects" *ColdSpring Harbor Laboratory Press* (1990), pp. 265-284.

Kato, et al.: "Purification and Characterization of DNA Topoisomerase IV in *Escherichia coli*" *Journal of Biological Chemistry* vol. 267, No. 3, pp. 25676-25684 (1992).

Hsieh, et al.: "Mechanistic Aspects of Type-II DNA Topoisomerase" *ColdSpring Harbor Laboratory Press* (1990), pp. 243-263.

Kim, et al.: "Identification of the Yeast TOP3 Gene Product as a Single Strand-specific DNA Topoisomerase" *Journal of Biological Chemistry*, vol. 267, No. 24, pp. 17178-17185 (1992).

Hanai, et al.: "Human TOP3: A single-copy gene encoding DNA topoisomerase III" *Proc. Natl. Acad. Sci. USA* vol. 93, pp. 3653-3657 (1996).

Siedlecki, et al.: "Characterization of a prokaryotic topoisomerase I activity in chloroplast extracts from spinach" *Nucleic Acids Research* vol. 11, pp. 1523-1536. (1983).

Kikuchi, et al.: "DNA Topology and Its Biological Effects" *ColdSpring Harbor Laboratory Press* (1990), pp. 285-298.

De La Tour, et al.: "Reverse Gyrase in Thermophilic Eubacteria" *Journal of Bacteriology* vol. 173, No. 12, pp. 3921-3923 (1991).

Slesarev, et al.: "DNA Topoisomerase III from Extremely Thermophilic Archaeobacteria" *Journal of Biological Chemistry* vol. 266, No. 19, pp. 12321-12328 (1991).

Neil Osheroff: "Biochemical Basis for the Interactions of Type I and Type II Topoisomerases with DNA" *Pharmac. Ther.* vol. 41, pp. 223-241 (1989).

Karl Drlica: "Biology of Bacterial Deoxyribonucleic Acid Topoisomerase" *Microbiological Reviews* vol. 48, No. 4, pp. 273-289 (1984).

Glisson, et al.: "DNA Topoisomerase II: A Primer on the Enzyme and its Unique Role as a Multidrug Target in Cancer Chemotherapy" *Pharmac. Ther.* vol. 32, pp. 89-106 (1987).

Bodley, et al.: "Topoisomerases as Novel Targets for Cancer Chemotherapy" *Bio/Technology* vol. 6 pp. 1315-1319 (1988).

Leroy F. Liu: "DNA Topoisomerase Poisons as Antitumor Drugs" *Annu. Rev. Biochem.* vol. 58, pp. 351-375 (1989).

Amy Luttinger: "The Twisted 'life' of DNA in the cell: bacterial topoisomerases" *Molecular Microbiology* vol. 15(4), pp. 601-606 (1995).

Berlani, et al.: "Transfer RNA genes in the cap-oxil region of yeast mitochondrial DNA" *Nucleic Acids Research* vol. 8, No. 21 (1980).

Joaquim Roca: "The mechanisms of DNA topoisomerases" *TIBS* vol. 20, pp. 156-160 (1995).

Furuta et al. *AIDS Research and Human Retroviruses*. 10: 1187-1188, 1994.

Primary Examiner—Carla J. Myers

Attorney, Agent, or Firm—Edward R. Gimmi; William T. King; Arthur E. Jackson

[57]

ABSTRACT

Topoisomerase III polypeptides and DNA and RNA encoding such Topoisomerase III polypeptides and a procedure for producing such polypeptides by recombinant techniques is disclosed. Also disclosed are methods for utilizing such Topoisomerase III for the treatment of infection, particularly bacterial infections. Antagonists against such Topoisomerase III and their use as a therapeutic to treat infections, particularly bacterial infections are also disclosed. Also disclosed are diagnostic assays for detecting diseases related to the presence of Topoisomerase III nucleic acid sequences and the polypeptides in a host. Also disclosed are diagnostic assays for detecting polynucleotides encoding Staphylococcal Topoisomerase III and for detecting the polypeptide in a host.

25 Claims, 3 Drawing Sheets

FIGURE 1.

TTCATTGTACTGTTGAGGAAGTTTATATAAGTATGATGCTGATTTCATAATTCGAATGTTCAATGAACGATTTTATTTGTGTAATATCATATAACTGAACTATGCTCATGTCATTACCTCCGTACTTTTTGTTACTTTTATTATATAGTATTTCAACTGAAATGAAAGTTAATAGTGATATTAACATGTTACAATACATTTAACACCATTTGAATTTAAATCAAAGATTAGTGGAATAGATGAAGCACGTTGAAATAAAAGAACGTATGAGAAAGGATAATTTATGAAATCTTTAATATTA
GCTGAAAAACCATCAGTTGCAAGAGATATTGCTGATGCTTTACAAATAAATCAGAAGCGT
AATGGTTACTTTGAAAAATAACCAATATATTGTCACGTGGGCGTTAGGTCATCTAGTGACA
AATGCGACACCTGAACAATACGATAAAAAATTTAAAGGAATGGCGATTAGAAGACCTTCCA
ATTATACCTAAATATATGAAAACTGTTGTTATTGGTAAAAACAAGCAAACAATTTAAAAACA
GTAAAAGCGTTAATTTTAGATAATAAAGTGAAAGATATTATTATTGCAACAGATGCTGGA
CGAGAAGGTGAAC TAGTTGCAAGATTGATTTTGGATAAAGTTGGTAACAAAAAGCCATC
CGTCGATTATGGATTAGCTCAGTTACTAAAAAAGCTATTCAACAAGGTTTTAAAAATTTA
AAAGACGGTCGTCAATATAACGATTTGTATTATGCAGCGTTAGCGAGAAGCGAGGCAGAT
TGGATTGTTGGGATTAATGCAACGCGTGCACTAACAACAAAGTATGATGCACAGCTATCC
CTGGGACGTGTTTCAGACACCAACGATTCAATTAGTAAATACACGACAACAAGAGATTAAT
CAGTTCAAACCACAACAATACTTTACATTATCATTAACGGTAAAAGGGTTTGATTTTCAG
CTAGAATCAAATCAGCGATATACCAATAAAGAACTTTAGAACAGATGGTTAATAATTTG
AAAAATGTCGATGGTAAGATTAAATCTGTTGCTACTAAACATAAGAAGTCGTATCCGCAA
TCACTGTACAATTTAACAGATTTACAACAAGATATGTATAGACGTTATAAAATTTGGACCT
AAAGAAACATTGAATACACTTCAAAGCTTATATGAGAGACATAAAGTCGTAACCTATCCA
AGAACAGATTCAAAC TATTTAACAAC TGATATGGTAGATACTATGAAAGAACGTATTCAG
GCGACGATGGCAACAACATATAAAGACCAAGCACGCCCATTAATGTCTAAAAACATTTTCA
TCAAAAATGTCGATATTTAATAATCAAAAAGTATCTGATC

Figure 1A

ACCATGCAATTATTCCCTACAGAAGTGAGACCTGTCATGTCAGACTTAAGTAATAGAGAAT
TAAAGTTATACGATATGATTGTCGAGCGTTTTTTAGAAAGCTTTAATGCCTCCGCACGAGT
ATGACGCGATAACTGTAACTTTAGAGGTTGCAGGGCACACATTTGTTTTGAAAGAGAATG
TAACAACGTGTTTTAGGTTTTAAATCTATTAGACAAGGTGAATCTATTACAGAGATGCAAC
AGCCTTTTTTCAGAAGGCGATGAAGTGAAGATTTCAAAAACAAACATTAGAGAACATGAAA
CAACACCTCCAGAATATTTTAATGAAGGTTTCGTTATTTAAAAGCGATGGAGAACCCTCAGA
ACTTTATTCAATTGAAGGATAAAAAATATGCGCAAACTTTAAAACAAACAGGTGGTATCG
GCACAGTTGCAACAAGGGCCGACATTATCGATAAATTATTTAATATGAATGCCATTGAAT
CAAGAGACGGTAAAATTAAAGTAACGTCAAAAGGTAAACAAATATTAGAATTAGCACCAG
AGAATTAACGTTCGCCACTTTTAACTGCACAATGGGAAGAAAAATTACTTTAATTGAAC
GTGGTAAATATCAGGCGAAAACATTTATTAATGAAATGAAAGATTTTACGAAAGATGTTG
TAAATGGGATTAAAAATAGTGATCGTAAATATAAACACGATAATTTAACAACCACAGAAT
GCCCAACGTGTGGTAAATTTCATGATTAAAGTTAAAACTAAAAATGGTCAGATGCTTGTGT
GCCAAGATCCATCTTGTAAAGACGAAAAAGAATGTACAGCGCAAAACAAATGCAAGATGTC
CAAACGTAAAAAGAAATTAACGTTGTTTGGTAAAGGGAAAGAAGCGGTATATCGTTGTG
TTTGTGGACATTCTGAAACGCAAGCACATATGGATCAGCGTATGAAGCTAAATCCTCTG
GTAAAGTATCTCGTAAAGAAATGAAAAAGTATATGAATAAAAAATGAAGGTTTAGACAATA
ATCCGTTTAAAGATGCATTAAAGAACTTGAATTTATAGATAAAATCGAACAAAGTTGAAT
CAGAAAAACGAAAAGTTCGCTTTTGGTATTGTTTTTTATTAAGAATGATATTAACTATT
AAGGTATTTTAAAAAAGGAGCATCCATTTCGTGAAAACTATTTCCAGTTCGATAAATAT
GGAACAACTTTAAAAGAGAAATCTTAGGCGGTATCACAACCTTCTTATCTATGGCCTAT
ATTTTAGCAGTTAACCCGCAAGTTTTAAGTTTAGCAGGTGTTAAAGGCGTATCAGAAGAT
ATGAAAATGGACCAAGGTGCCATTTTTGTAGCGACTGCATTAGCAGCATTTGTAGGCTCG
CTATTCATGGGACTAATAGCTAAATATCCAATCGCATTAGCACCAGGTATGGGATTGGAA
TTC

FIGURE 2.

MKSLILA EKPSVARDIADALQINQKRNGYFENNQYIVTWALGHLVTNATPEQYDKNLKEW
RLEDLP IIPKYMKT VVIGKTSKQFKTVKALILDNKVKD I I IATDAGREGELVARLIILDKV
GNKKPIRRLWISSVT KKA IQQGFKNLKDGRQYNDLYYAALARSEADWIVGINATRALTTK
YDAQLSLGRVQTPTIQLVNTRQQEINQFKPQQYFTLSLTVKGFDQLESNQRYTNKETLE
QMVNNLKNVDGKIKSVATKHKKSYPQSLYNLTDLQQDMYRRYKIGPKETLNTLQSLYERH
KVVTPRTDSNYLT TDMVD TMKERIQATMATTYKDQARPLMSKTFSSKMSIFNNQKVSDH
HAIIPTEVRPVMSDLSNRELKLYDMIVERFLEALMPPEYDAITVTLEVAGHTFVLKENV
TTVLGFKSIRQGESITEMQQPFSEGDEVKISKTNIREHETTPPEYFNEGSLLKAMENPQN
FIQLKDKKYAQT LKQTGGIGTVATRAD I IDKLFNMNAIESRDGKIKVTSKGKQILELAPE
ELTSPLLTAQWEEKLLLIERGKYQAKTFINEMKDFTKD VVNGIKNSDRKYKHDNLTTTEC
PTCGKFM I KVKTKNGQMLVCQDPSC TKKNVQRKTNARCPNCKKKLTLFGKGKEAVYRCV
CGHSETQAHMDQRMKSKSSGKVS RKEMKKYMNKNEGLDNNPFKDALKNLNL

TOPOISOMERASE III

RELATED APPLICATIONS

This Application claims benefit of U.S. Provisional Application Ser. No. 60/028,417, filed Oct. 15, 1996.

FIELD OF THE INVENTION

This invention relates, in part, to newly identified polynucleotides and polypeptides; variants and derivatives of these polynucleotides and polypeptides; processes for making these polynucleotides and these polypeptides, and their variants and derivatives; agonists and antagonists of the polypeptides; and uses of these polynucleotides, polypeptides, variants, derivatives, agonists and antagonists. In particular, in these and in other regards, the invention relates to polynucleotides and polypeptides of bacterial "Topoisomerase III".

BACKGROUND OF THE INVENTION

Among the more effective antibiotics are those that interfere with common modes of bacterial gene expression, regulation or activity. Recently, the supercoiling of DNA had been suggested as a possible mode of virulence gene regulation. Local increases or decreases in DNA density, due to supercoiling, have been associated with responses to various environmental conditions such as, temperature, anaerobiosis, and osmolarity. Appropriate regulation of the accessibility of groups of genes to components of the transcriptional apparatus by increasing or decreasing supercoiling of spatially organized genes may represent an infecting pathogen's effective response to such environmental conditions. Enzymes, such as DNA topoisomerases including type 1 topoisomerases and DNA gyrases, have been identified which function to effect the levels of DNA supercoiling. Such enzymes represent useful targets against which to screen compounds as potential antibiotics.

DNA transformations performed by DNA topoisomerases are accomplished by the cleavage of either a single strand or both strands. The unit change in the Linking number (Lk) resulting from such transformations is the best operational distinction between the two classes of topoisomerases (P. O. Brown & N. R. Cozzarelli, *Science* 206:1081-1083 (1979)). The linking number (Lk) is the algebraic number of times one strand crosses the surface stretched over the other strand. DNA topoisomerases whose reactions proceed via a transient single-stranded break and changing the Lk in steps of one are classified as type 1, while enzymes whose reactions proceed via double-stranded breaks and changing the Lk in steps of two are classified as type 2.

Members of type 2 topoisomerase family include DNA gyrase, bacterial DNA topoisomerase IV, T-even phage DNA topoisomerases, eukaryotic DNA topoisomerase II, and thermophilic topoisomerase II from *Sulfolobus acidocaldarius* (see: A. Kikuchi et al., *Syst. Appl. Microbiol.* 7: 72-78 (1986); J. Kato et al., *J. Biol. Chem.* 267: 25676-25684 (1992); W. M. Huang in *DNA Topology and Its Biological Effects* (N. R. Cozzarelli and J. C. Wang, eds., Cold Spring Harbor Laboratory Press, New York, 1990), pp. 265-284; T.-S. Hsieh in *DNA Topology and Its Biological Effects* (N. R. Cozzarelli and J. C. Wang, eds., Cold Spring Harbor Laboratory Press, New York, (1990), pp. 243-263)). The coding sequences of a dozen or so type 2 enzymes have been determined, and the data suggest that all these enzymes are evolutionary and structurally related. Topological reactions catalyzed by type 2 topoisomerases include introduction of

negative supercoils into DNA (DNA gyrase), relaxation of supercoiled DNA, catenation (or decatenation) of duplex circles, knotting and unknotting of DNA.

The family of type 1 topoisomerases comprises bacterial topoisomerase I, *E. coli* topoisomerase III, *S. cerevisiae* topoisomerase III (R. A. Kim & J. C. Wang, *J. Biol. Chem.* 267: 17178-17185 (1992), human topoisomerase III (Hanai et al., *Proc. Natl. Acad. Sci.* 93:3653-3657 (1996)), the type 1 topoisomerase from chloroplasts that closely resembles bacterial enzymes (J. Siedlecki et al., *Nucleic Acids Res.* 11: 1523-1536 (1983), thermophilic reverse gyrases (A. Kikuchi, In *DNA: "Topology and Its Biological Effects"* (N. R. Cozzarelli and J. C. Wang, eds., Cold Spring Harbor Laboratory Press, New York, 1990, pp. 285-298); C. Bouthier de la Tour et al., *J. Bact.* 173: 3921-3923 (1991), thermophilic *D. amylophilus* topoisomerase III (A. I. Slesarev et al., *J. Biol. Chem.* 266: 12321-12328 (1991), nuclear topoisomerases I and closely related enzymes from mitochondria and poxviruses (N. Osheroff, *Pharmac. Ther.* 41: 223-241 (1989)). With respect to the mechanism of catalysis these topoisomerases can be divided into two groups. Group A consists of enzymes that require a divalent cation for activity, and form a transient covalent complex with the 5'-phosphoryl termini (prokaryotic type 1 topoisomerases, *S. cerevisiae* topoisomerase III, and human topoisomerase III). Group B includes type 1 topoisomerases that do not require a divalent cation for activity, and bind covalently to the 3'-phosphoryl termini (nuclear topoisomerases I, enzymes from mitochondria and poxviruses commonly called eukaryotic topoisomerases I). Type 1 topoisomerases can carry out the following topological reactions: they relax supercoiled DNA (except of reverse gyrases), catenate (or decatenate) single-stranded circular DNAs or duplexes providing that at least one of the molecules contains a nick or gap, or interact with single-stranded circles to introduce topological knots (type 1-group A topoisomerases). Reverse gyrase, belonging to type 1-group A topoisomerases, is the only topoisomerase shown to be able to introduce positive supercoils into cDNA.

Research on DNA topoisomerases has progressed from DNA enzymology to developmental therapeutics. Bacterial DNA topoisomerase II is an important therapeutic target of quinolone antibiotics; mammalian DNA topoisomerase II is the cellular target of many potent antitumor drugs (K. Drlica, *Microbiol. Rev.* 48: 273-289 (1984) and *Biochemistry* 27: 2253-2259 (1988); B. S. Glisson & W. E. Ross, *Pharmacol. Ther.* 32: 89-106 (1987); A. L. Bodley & L. F. Liu, *Biotechnology* 6: 1315-1319 (1988); L. F. Liu, *Annu. Rev. Biochem.* 58: 351-375 (1989)). These drugs, referred to as topoisomerase II poisons, interfere with the breakage-rejoining reaction of type II topoisomerase by trapping a key covalent reaction intermediate, termed the cleavable complex. Mammalian topoisomerase I is the cellular target of the antitumor drug topotecan (U.S. Pat. No. 5,004,758), which also traps the covalent reaction intermediate.

As mentioned above, bacterial type I topoisomerases (topoisomerase I & III) are enzymes that alter DNA topology and are involved in a number of crucial cellular processes including replication, transcription and recombination (Luttinger, A., *Molecular Microbiol.* 15(4): 601-608 (1995)). These enzymes act by transiently breaking one strand of DNA, passing a single or double strand of DNA through the break and finally resealing the break. Cleavage of the DNA substrate forms a covalent linkage between a tyrosine residue of the enzyme and the 5' end of the DNA chain at the cleavage site (Roca, J. A., *TIBS* 20:156-160 (1995)).

Enzyme inhibition which leads to the stabilization of the covalent-enzyme-DNA complex (cleavable complex), will invoke chromosomal damage, and bacterial cell death. Furthermore, this mechanism has the potential of leading to cell death by virtue of a single inhibition event. A small molecular weight inhibitor, which acts by stabilization of the cleavable complex may act on both topoisomerase I and III because of the extensive amino acid sequence similarity between them, particularly in the region of their active sites. The likelihood of future high level resistance to such agents arising from point mutation may therefore be low.

Inhibitors of type I topoisomerases, for example, those able to stabilize the protein in a covalent complex with DNA would be lethal or inhibitory to the bacterium and thereby have utility in anti-bacterial therapy. It is particularly preferred to employ *Staphylococcal* genes and gene products as targets for the development of antibiotics. The *Staphylococci* make up a medically important genera of microbes. They are known to produce two types of disease, invasive and toxigenic. Invasive infections are characterized generally by abscess formation effecting both skin surfaces and deep tissues. *S. aureus* is the second leading cause of bacteremia in cancer patients. Osteomyelitis, septic arthritis, septic thrombophlebitis and acute bacterial endocarditis are also relatively common. There are at least three clinical conditions resulting from the toxigenic properties of *Staphylococci*. The manifestation of these diseases result from the actions of exotoxins as opposed to tissue invasion and bacteremia. These conditions include: *Staphylococcal* food poisoning, scalded skin syndrome and toxic shock syndrome.

SUMMARY OF THE INVENTION

Toward these ends, and others, it is an object of the present invention to provide polypeptides, inter alia, that have been identified as novel Topoisomerase III by homology between the amino acid sequence set out in FIG. 2 (SEQ ID NO: 2) and known amino acid sequences of other proteins such as *Haemophilus influenzae* topoisomerase III.

It is a further object of the invention, moreover, to provide polynucleotides that encode Topoisomerase III, particularly polynucleotides that encode the polypeptide herein designated bacterial Topoisomerase III.

In a particularly preferred embodiment of this aspect of the invention the polynucleotide comprises the region encoding Topoisomerase III in the sequence set out in FIG. 1 (SEQ ID NO: 1).

In another particularly preferred embodiment of the present invention there is a novel Topoisomerase III protein from *Staphylococcus aureus* comprising the amino acid sequence of (SEQ ID NO: 2), or a fragment, analogue or derivative thereof.

In accordance with this aspect of the present invention there is provided an isolated nucleic acid molecule encoding a mature polypeptide expressible by the *Staphylococcus aureus* polynucleotide contained in deposited strain NCIMB 40771.

In accordance with this aspect of the invention there are provided isolated nucleic acid molecules encoding Topoisomerase III, particularly *Staphylococcal* Topoisomerase III, including mRNAs, cDNAs, genomic DNAs and, in further embodiments of this aspect of the invention, biologically, diagnostically, clinically or therapeutically useful variants, analogs or derivatives thereof, or fragments thereof, including fragments of the variants, analogs and derivatives.

Among the particularly preferred embodiments of this aspect of the invention are naturally occurring allelic variants of Topoisomerase III.

In accordance with this aspect of the invention there are provided novel polypeptides of *Staphylococcal* origin referred to herein as Topoisomerase III as well as biologically, diagnostically or therapeutically useful fragments thereof, as well as variants, derivatives and analogs of the foregoing and fragments thereof.

It also is an object of the invention to provide Topoisomerase III polypeptides, particularly bacterial Topoisomerase III polypeptides, that may be employed for therapeutic purposes, for example, to treat disease, including treatment by conferring host immunity against bacterial infections, such as *Staphylococcal* infections.

In accordance with yet a further aspect of the present invention, there is provided the use of a polypeptide of the invention, in particular a fragment thereof, for therapeutic or prophylactic purposes, for example, as an antibacterial agent or a vaccine.

In accordance with another aspect of the present invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunization.

Among the particularly preferred embodiments of this aspect of the invention are variants of Topoisomerase III polypeptide encoded by naturally occurring alleles of the Topoisomerase III gene.

It is another object of the invention to provide a process for producing the aforementioned polypeptides, polypeptide fragments, variants and derivatives, fragments of the variants and derivatives, and analogs of the foregoing.

In a preferred embodiment of this aspect of the invention there are provided methods for producing the aforementioned Topoisomerase III polypeptides comprising culturing host cells having expressibly incorporated therein an exogenously-derived Topoisomerase III-encoding polynucleotide under conditions for expression of Topoisomerase III in the host and then recovering the expressed polypeptide.

In accordance with another object the invention there are provided products, compositions, processes and methods that utilize the aforementioned polypeptides and polynucleotides, inter alia, for research, biological, clinical and therapeutic purposes.

In accordance with yet another aspect of the present invention, there are provided inhibitors of such polypeptides, useful as antibacterial agents. In particular, there are provided antibodies against such polypeptides.

In accordance with certain preferred embodiments of this and other aspects of the invention there are probes that hybridize to bacterial Topoisomerase III sequences useful for detection of bacterial infection.

In certain additional preferred embodiments of this aspect of the invention there are provided antibodies against Topoisomerase III polypeptides. In certain particularly preferred embodiments in this regard, the antibodies are selective for *Staphylococcal* Topoisomerase III.

In accordance with another aspect of the present invention, there are provided Topoisomerase III agonists. Among preferred agonists are molecules that mimic Topoisomerase III, that bind to Topoisomerase III-binding molecules, and that elicit or augment Topoisomerase III-induced responses. Also among preferred agonists are molecules that interact with Topoisomerase III encoding genes or Topoisomerase III polypeptides, or with other modulators of Topoisomerase III activities, and thereby potentiate or augment an effect of Topoisomerase III or more than one

effect of Topoisomerase III and which are also preferably bacteriostatic or bactericidal.

In accordance with yet another aspect of the present invention, there are provided Topoisomerase III antagonists. Among preferred antagonists are those which bind to Topoisomerase III so as to inhibit the binding of Topoisomerase III-binding molecules or to stabilize the complex formed between Topoisomerase III and Topoisomerase III binding molecule to prevent further biological activity arising from the Topoisomerase III. Also among preferred antagonists are molecules that bind to or interact with Topoisomerase III so as to inhibit an effect of Topoisomerase III or more than one effect of Topoisomerase III or which prevent expression of Topoisomerase III and which are also preferably bacteriostatic or bactericidal.

In a further aspect of the invention there are provided compositions comprising a Topoisomerase III polynucleotide or a Topoisomerase III polypeptide for administration to cells in vitro, to cells ex vivo and to cells in vivo, or to a multicellular organism. In certain preferred embodiments of this aspect of the invention, the compositions comprise a Topoisomerase III polynucleotide for expression of a Topoisomerase III polypeptide in a host organism to raise an immunological response, preferably to raise immunity in such host against Staphylococci or related organisms.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill from the following description. It should be understood, however, that the following description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following description and from reading the other parts of the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings depict certain embodiments of the invention. They are illustrative only and do not limit the invention otherwise disclosed herein.

FIG. 1 shows a polynucleotide sequence of *Staphylococcus aureus* Topoisomerase III, comprising a sequence of *Staphylococcus aureus* Topoisomerase III gene and surrounding area (coding sequence underlined) (SEQ ID NO: 1).

FIG. 2 shows the amino acid sequence of *Staphylococcus aureus* Topoisomerase III (SEQ ID NO: 2) deduced from the polynucleotide coding sequence of FIG. 1 (SEQ ID NO: 1).

GLOSSARY

The following illustrative explanations are provided to facilitate understanding of certain terms used frequently herein, particularly in the Examples. The explanations are provided as a convenience and are not limitative of the invention.

Topoisomerase III-BINDING MOLECULE, as used herein, refers to molecules or ions which bind or interact specifically with Topoisomerase in polypeptides or polynucleotides of the present invention, including, for example enzyme substrates, such as supercoiled DNA, cell membrane components and classical receptors. Binding between polypeptides of the invention and such molecules, including binding or interaction molecules may be exclusive to polypeptides of the invention, which is preferred, or it may be highly specific for polypeptides of the invention, which

is also preferred, or it may be highly specific to a group of proteins that includes polypeptides of the invention, which is preferred, or it may be specific to several groups of proteins at least one of which includes a polypeptide of the invention. Binding molecules also include antibodies and antibody-derived reagents that bind specifically to polypeptides of the invention.

GENETIC ELEMENT generally means a polynucleotide comprising a region that encodes a polypeptide or a polynucleotide region that regulates replication, transcription or translation or other processes important to expression of the polypeptide in a host cell, or a polynucleotide comprising both a region that encodes a polypeptide and a region operably linked thereto that regulates expression. Genetic elements may be comprised within a vector that replicates as an episomal element; that is, as a molecule physically independent of the host cell genome. They may be comprised within plasmids. Genetic elements also may be comprised within a host cell genome; not in their natural state but, rather, following manipulation such as isolation, cloning and introduction into a host cell in the form of purified DNA or in a vector, among others.

HOST CELL is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous polynucleotide sequence.

IDENTITY or SIMILARITY, as known in the art, are relationships between two polypeptide sequences or two polynucleotide sequences, as determined by comparing the sequences. In the art, identity also means the degree of sequence relatedness between two polypeptide or two polynucleotide sequences as determined by the match between two strings of such sequences. Both identity and similarity can be readily calculated (*Computational Molecular Biology*, Lesk, A. M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D. W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data*, Part I, Griffin, A. M., and Griffin, H. G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity and similarity between two polynucleotide or two polypeptide sequences, both terms are well known to skilled artisans (*Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991; and Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48: 1073 (1988). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to those disclosed in Carillo, H., and Lipman, D., *SIAM J. Applied Math.*, 48:1073 (1988). Preferred methods to determine identity are designed to give the largest match between the two sequences tested. Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., et al., *Nucleic Acids Research* 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S. F. et al., *J. Mol. Biol.* 215: 403 (1990)).

ISOLATED means altered "by the hand of man" from its natural state; i.e., that, if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a naturally occurring polynucleotide or a polypeptide naturally present in a living organism in its natural state is not "isolated," but the same polynucleotide or

polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. For example, with respect to polynucleotides, the term isolated means that it is separated from the chromosome and cell in which it naturally occurs. As part of or following isolation, such polynucleotides can be joined to other polynucleotides, such as DNAs, for mutagenesis, to form fusion proteins, and for propagation or expression in a host, for instance. The isolated polynucleotides, alone or joined to other polynucleotides such as vectors, can be introduced into host cells, in culture or in whole organisms. Introduced into host cells in culture or in whole organisms, such DNAs still would be isolated, as the term is used herein, because they would not be in their naturally occurring form or environment. Similarly, the polynucleotides and polypeptides may occur in a composition, such as a media formulations, solutions for introduction of polynucleotides or polypeptides, for example, into cells, compositions or solutions for chemical or enzymatic reactions, for instance, which are not naturally occurring compositions, and, therein remain isolated polynucleotides or polypeptides within the meaning of that term as it is employed herein.

POLYNUCLEOTIDE(S) generally refers to any polyribonucleotide or polynucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. Thus, for instance, polynucleotides as used herein refers to, among others, single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions or single-, double- and triple-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded, or triple-stranded, or a mixture of single- and double-stranded regions. In addition, polynucleotide as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide. As used herein, the term polynucleotide includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including inter alia, simple and complex cells. The term polynucleotide(s) embrace short polynucleotides often referred as oligonucleotides.

POLYPEPTIDES, as used herein, includes all polypeptides as described below. The basic structure of polypeptides is well known and has been described in innumerable textbooks and other publications in the art. In this context, the term is used herein to refer to any peptide or protein comprising two or more amino acids joined to each other in a linear chain by peptide bonds. As used herein, the term refers to both short chains, which also commonly are referred to in the art as peptides, oligopeptides and

oligomers, for example, and to longer chains, which generally are referred to in the art as proteins, of which there are many types. It will be appreciated that polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids, and that many amino acids, including the terminal amino acids, may be modified in a given polypeptide, either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques which are well known to the art. Even the common modifications that occur naturally in polypeptides are too numerous to list exhaustively here, but they are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art. Among the known modifications which may be present in polypeptides of the present are, to name an illustrative few, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Such modifications are well known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance *PROTEINS—STRUCTURE AND MOLECULAR PROPERTIES*, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, F., *Posttranslational Protein Modifications: Perspectives and Prospects*, pgs. 1–12 in *POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS*, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., *Meth. Enzymol.* 182:626–646 (1990) and Rattan et al., *Protein Synthesis: Posttranslational Modifications and Aging*, *Ann. N.Y. Acad. Sci.* 663: 48–62 (1992). It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli* or other cells, prior to proteolytic processing, almost invariably will be N-formylmethionine. During post-translational modifi-

cation of the peptide, a methionine residue at the NH₂-terminus may be deleted. Accordingly, this invention contemplates the use of both the methionine-containing and the methionineless amino terminal variants of the protein of the invention. The modifications that occur in a polypeptide often will be a function of how it is made. For polypeptides made by expressing a cloned gene in a host, for instance, the nature and extent of the modifications in large part will be determined by the host cell posttranslational modification capacity and the modification signals present in the polypeptide amino acid sequence. For instance, as is well known, glycosylation often does not occur in bacterial hosts such as, for example, *E. coli*. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell. Insect cells often carry out the same posttranslational glycosylations as do mammalian cells and, for this reason, insect cell expression systems have been developed to express efficiently mammalian proteins having native patterns of glycosylation. Similar considerations apply to other modifications. It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. In general, as used herein, the term polypeptide encompasses all such modifications, particularly those that are present in polypeptides synthesized by expressing a polynucleotide in a host cell.

VARIANT(S) of polynucleotides or polypeptides, as the term is used herein, are polynucleotides or polypeptides that differ from a reference polynucleotide or polypeptide, respectively. Variants in this sense are described below and elsewhere in the present disclosure in greater detail. With reference to polynucleotides, generally, differences are limited such that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical. As noted below, changes in the nucleotide sequence of the variant may be silent. That is, they may not alter the amino acids encoded by the polynucleotide. Where alterations are limited to silent changes of this type, a variant will encode a polypeptide with the same amino acid sequence as the reference. Also as noted below, changes in the nucleotide sequence of the variant may alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Such nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. With reference to polypeptides generally, differences are limited so that the sequences of the reference and the variant are closely similar overall and, in many region, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

DESCRIPTION OF THE INVENTION

The present invention relates to novel Topoisomerase III polypeptides and polynucleotides encoding same, among other things, as described in greater detail below. In particular, the invention relates to polypeptides and polynucleotides of a novel Topoisomerase III gene of *Staphylococcus aureus*, which is related by amino acid sequence homology to, for example, *Haemophilus influenzae* topoisomerase III protein. The closest relatives to *Staphylococcus* topoisomerase III are the *Haemophilus influenzae* topoisomerase III, which is 33% identical and 52% similar at the amino acid level, and 57% identical at the nucleotide level, and *E. coli* topoisomerase III, which is 30% identical and

53% similar at the amino acid level, and 66% identical at the nucleotide level. These homology determinations were made using the Genetics Computer Group BESTFIT program.

The invention relates especially to *Staphylococcal* Topoisomerase III having the nucleotide and amino acid sequences set out in FIG. 1 (SEQ ID NO: 1) and FIG. 2 (SEQ ID NO: 2), and to the Topoisomerase III nucleotide and amino acid sequences of the DNA isolatable from Deposit No. NC1MB 40771, which is herein referred to as "the deposited organism" or as the "DNA of the deposited organism." It will be appreciated that the nucleotide and amino acid sequences set out in FIG. 1 (SEQ ID NO: 1) and FIG. 2 (SEQ ID NO: 2) were obtained by sequencing the DNA of the deposited organism. Hence, the sequence of the deposited clone is controlling as to any discrepancies between it (and the sequence it encodes) and the sequences of FIG. 1 (SEQ ID NO: 1) and FIG. 2 (SEQ ID NO: 2).
Polynucleotides

In accordance with one aspect of the present invention, there are provided isolated polynucleotides which encode the *Staphylococcal* Topoisomerase III polypeptide having the deduced amino acid sequence of FIG. 2 (SEQ ID NO: 2).

Using the information provided herein, such as the polynucleotide sequence set out in FIG. 1 (SEQ ID NO: 1), a polynucleotide of the present invention encoding Topoisomerase III polypeptide may be obtained using standard cloning and screening procedures. To obtain the polynucleotide encoding the protein using the DNA sequence given in SEQ ID NO: 1 typically a library of clones of chromosomal DNA of *S. aureus* WCUH 29 in *E. coli* or some other suitable host is probed with a radiolabelled oligonucleotide, preferably a 17 mer or longer, derived from the sequence of FIG. 1. Clones carrying DNA identical to that of the probe can then be distinguished using high stringency washes. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E. F. and Sambrook, J. in MOLECULAR CLONING, A Laboratory Manual (2nd edition 1989 Cold Spring Harbor Laboratory. see Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70). Illustrative of the invention, the polynucleotide set out in FIG. 1 (SEQ ID NO: 1) was discovered in a DNA library derived from *Staphylococcus aureus* NC1MB 40771 as described in Example I.

Topoisomerase III of the invention is structurally related to other proteins of the bacterial Topoisomerase III family, as shown by comparing the sequence encoding Topoisomerase III from the deposited clone with that of sequence reported in the literature. A preferred DNA sequence is set out in FIG. 1 (SEQ ID NO: 1). It contains an open reading frame encoding a protein of about 711 amino acid residues. The protein exhibits greatest homology to *Haemophilus influenzae* topoisomerase m protein among known proteins.

Polynucleotides of the present invention may be in the form of RNA, such as mRNA, or in the form of DNA, including, for instance, cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The DNA may be double-stranded or single-stranded. Single-stranded DNA may be the coding strand, also known as the sense strand, or it may be the non-coding strand, also referred to as the anti-sense strand.

The coding sequence which encodes the polypeptide may be identical to the coding sequence of the polynucleotide shown in FIG. 1 (SEQ ID NO: 1). It also may be a polynucleotide with a different sequence, which, as a result of the redundancy (degeneracy) of the genetic code, encodes the polypeptide of FIG. 2 (SEQ ID NO: 2).

Polynucleotides of the present invention which encode the polypeptide of FIG. 2 (SEQ ID NO: 2) may include, but are not limited to the coding sequence for the mature polypeptide, by itself; the coding sequence for the mature polypeptide and additional coding sequences, such as those encoding a leader or secretory sequence, such as a pre-, or pro- or prepro- protein sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with additional, non-coding sequences, including for example, but not limited to non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription (including termination signals, for example), ribosome binding, mRNA stability elements, and additional coding sequence which encode additional amino acids, such as those which provide additional functionalities. The DNA may also comprise a promoter region which functions to direct the transcription of the mRNA encoding the Topoisomerase III of this invention. Such promoter may be independently useful to direct the transcription of heterologous gene in recombinant expression system. Furthermore, the polypeptide may be fused to a marker sequence, such as a peptide, which facilitates purification of the fused polypeptide. In certain embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, such as the tag provided in the pQE vector (Qiagen, Inc.), among others, many of which are commercially available. As described in Gentz et al., *Proc. Nat'l. Acad. Sci., USA* 86: 821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. The HA tag may also be used to create fusion proteins and corresponds to an epitope derived of influenza hemagglutinin protein, which has been described by Wilson et al., *Cell* 37: 767 (1984), for instance.

In accordance with the foregoing, the term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides which include a sequence encoding a polypeptide of the present invention, particularly bacterial, and more particularly *Staphylococcus aureus* Topoisomerase III having the amino acid sequence set out in FIG. 2 (SEQ ID NO: 2). The term encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by integrated phage or insertion sequence or editing) together with additional regions, that also may contain coding and/or non-coding sequences.

The present invention further relates to variants of the herein above described polynucleotides which encode for fragments, analogs and derivatives of the polypeptide having the deduced amino acid sequence of FIG. 2 (SEQ ID NO: 2). A variant of the polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis techniques, including those applied to polynucleotides, cells or organisms.

Among variants in this regard are variants that differ from the aforementioned polynucleotides by nucleotide substitutions, deletions or additions. The substitutions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions.

Among the particularly preferred embodiments of the invention in this regard are polynucleotides encoding polypeptides having the amino acid sequence of *Staphylococcal* Topoisomerase III set out in FIG. 2 (SEQ ID NO: 2); variants, analogs, derivatives and fragments thereof.

Further particularly preferred in this regard are polynucleotides encoding Topoisomerase III variants, analogs, derivatives and fragments, and variants, analogs and derivatives of the fragments, which have the amino acid sequence of *Staphylococcal* Topoisomerase III polypeptide of FIG. 2 (SEQ ID NO: 2) in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of Topoisomerase III. Also especially preferred in this regard are conservative substitutions. Most highly preferred are polynucleotides encoding polypeptides having the amino acid sequence of FIG. 2 (SEQ ID NO: 2), without substitutions.

Further preferred embodiments of the invention are polynucleotides that are at least 70% identical to a polynucleotide encoding Topoisomerase III polypeptide having the amino acid sequence set out in FIG. 2 (SEQ ID NO: 2), and polynucleotides which are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80% identical to a polynucleotide encoding Topoisomerase III polypeptide of the *Staphylococcus aureus* DNA of the deposited clone and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical to the same are particularly preferred, and among these particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

Particularly preferred embodiments in this respect, moreover, are polynucleotides which encode polypeptides which retain substantially the same biological function or activity as the mature polypeptide encoded by the DNA of FIG. 1 (SEQ ID NO: 1).

The present invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the present invention especially relates to polynucleotides which hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for RNA, cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding Topoisomerase III and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to the Topoisomerase III gene. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

For example, the coding region of the Topoisomerase III gene may be isolated by screening using the known DNA sequence to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the present invention is then used to screen a

library of cDNA, genomic DNA or mRNA to determine which members of the library to which the probe hybridizes.

The polynucleotides and polypeptides of the present invention may be employed as research reagents and materials for discovery of treatments of and diagnostics for disease, particularly human disease, as further discussed herein relating to polynucleotide assays.

The polynucleotides of the invention that are oligonucleotides, derived from the sequences (SEQ ID NO: 1) may be used as PCR primers in the process herein described to determine whether or not the *Staphylococcus aureus* genes identified herein in whole or in part are transcribed in infected tissue. It is recognized that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

The polynucleotides may encode a polypeptide which is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may allow protein transport, may lengthen or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case, in vivo, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

In sum, a polynucleotide of the present invention may encode a mature protein, a mature protein plus a leader sequence (which may be referred to as a preprotein), a precursor of a mature protein having one or more prosequences which are not the leader sequences of a preprotein, or a preproprotein, which is a precursor to a proprotein, having a leader sequence and one or more prosequences, which generally are removed during processing steps that produce active and mature forms of the polypeptide.

Deposited materials

Staphylococcus aureus WCUH 29 was deposited at the National Collection of Industrial and Marine Bacteria Ltd. (NCIMB), 23 St. Machar Drive, Aberdeen, Scotland under number NCIMB 40771 on Sep. 11, 1995.

The deposit has been made under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for Purposes of Patent Procedure. The strain will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposit is provided merely as convenience to those of skill in the art and is not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. §112.

The sequence of the polynucleotides contained in the deposited material, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

Polypeptides

The present invention further relates to a bacterial Topoisomerase III polypeptide that has the deduced amino acid sequence of FIG. 2 (SEQ ID NO: 2).

The invention also relates to fragments, analogs and derivatives of these polypeptides. The terms "fragment," "derivative" and "analog" when referring to the polypeptide

of FIG. 2 (SEQ ID NO: 2), means a polypeptide which retains essentially the same biological function or activity as such polypeptide. Fragments derivatives and analogs that retain at least 90% of the activity of the native Topoisomerase III are preferred. Fragments derivatives and analogs that retain at least 95% of the activity of the native Topoisomerase III are preferred. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide.

The polypeptide of the present invention may be a recombinant polypeptide, a natural polypeptide or a synthetic polypeptide. In certain preferred embodiments it is a recombinant polypeptide.

The fragment, derivative or analog of the polypeptide of FIG. 2 (SEQ ID NO: 2) may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be obtained by those of ordinary skill in the art, from the teachings herein.

Among the particularly preferred embodiments of the invention in this regard are polypeptides having the amino acid sequence of *Staphylococcus aureus* Topoisomerase III set out in FIG. 2 (SEQ ID NO: 2), variants, analogs, derivatives and fragments thereof, and variants, analogs and derivatives of the fragments.

Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Further particularly preferred in this regard are variants, analogs, derivatives and fragments, and variants, analogs and derivatives of the fragments, having the amino acid sequence of the Topoisomerase III polypeptide of FIG. 2 (SEQ ID NO: 2), in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the Topoisomerase III. Also especially preferred in this regard are conservative substitutions. Most highly preferred are polypeptides having the amino acid sequence of FIG. 2 (SEQ ID NO: 2) without substitutions.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The polypeptides of the present invention include the polypeptide of FIG. 2 (SEQ ID NO: 2), in particular the mature polypeptide as well as polypeptides which have at least 80% identity to the polypeptide of FIG. 2 (SEQ ID NO: 2).

2) and preferably at least 90% similarity (more preferably at least 90% identity) to the polypeptide of FIG. 2 (SEQ ID NO: 2) and more preferably at least 95% similarity; and still more preferably at least 95% identity to the polypeptide of FIG. 2 (SEQ ID NO: 2) and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 contiguous amino acids and more preferably at least 50 contiguous amino acids.

Fragments or portions of the polypeptides of the present invention may be employed for producing the corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

Fragments

Also among preferred embodiments of this aspect of the present invention are polypeptides comprising fragments of Topoisomerase III, most particularly fragments of Topoisomerase III having the amino acid set out in FIG. 2 (SEQ ID NO: 2), and fragments of variants and derivatives of the Topoisomerase III of FIG. 2 (SEQ ID NO: 2).

In this regard a fragment is a polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned Topoisomerase III polypeptides and variants or derivatives thereof.

Such fragments may be "free-standing," i.e., not part of or fused to other amino acids or polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the presently discussed fragments most preferably form a single continuous region. However, several fragments may be comprised within a single larger polypeptide. For instance, certain preferred embodiments relate to a fragment of a Topoisomerase III polypeptide of the present invention comprised within a precursor polypeptide designed for expression in a host and having heterologous pre and pro-polypeptide regions fused to the amino terminus of the Topoisomerase III fragment and an additional region fused to the carboxyl terminus of the fragment. Therefore, fragments in one aspect of the meaning intended herein, refers to the portion or portions of a fusion polypeptide or fusion protein derived from Topoisomerase III.

Representative examples of polypeptide fragments of the invention, include, for example, may be mentioned those which have from about 5-15, 10-20, 15-40, 30-55, 41-75, 41-80, 41-90, 50-100, 75-100, 90-115, 100-125, and 110-140, 120-150, 200-300, 1-175 or 1-711 amino acids long.

In this context about includes the particularly recited range and ranges larger or smaller by several, a few, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes.

Among especially preferred fragments of the invention are truncation mutants of Topoisomerase III. Truncation mutants include Topoisomerase III polypeptides having the amino acid sequence of FIG. 2 (SEQ ID NO: 2), or of variants or derivatives thereof, except for deletion of a continuous series of residues (that is, a continuous region, part or portion) that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus or, as in double truncation mutants, deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Fragments having the size ranges set out above also are preferred embodiments of truncation fragments, which are especially

preferred among fragments generally. Degradation forms of the polypeptides of the invention in a host cell are also preferred.

Also preferred in this aspect of the invention are fragments characterized by structural or functional attributes of Topoisomerase III. Preferred embodiments of the invention in this regard include fragments that comprise alpha-helix and alpha-helix forming regions ("alpha-regions"), beta-sheet and beta-sheet-forming regions ("beta-regions"), turn and turn-forming regions ("turn-regions"), coil and coil-forming regions ("coil-regions"), hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions and high antigenic index regions of Topoisomerase III.

Further preferred regions are those that mediate activities of Topoisomerase III. Most highly preferred in this regard are fragments that have a chemical, biological or other activity of Topoisomerase III, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Routinely one generates the fragment by well-known methods then compares the activity of the fragment to native Topoisomerase I in a convenient assay such as listed hereinbelow. Highly preferred in this regard are fragments that contain regions that are homologs in sequence, or in position, or in both sequence and to active regions of related polypeptides, such as the related polypeptides set out in FIG. 2 (SEQ ID NO: 2), which include *E. coli* topoisomerase III and *H. influenzae* topoisomerase III. Among particularly preferred fragments in these regards are truncation mutants, as discussed above. Further preferred polynucleotide fragments are those that are antigenic or immunogenic in an animal, especially in a human.

It will be appreciated that the invention also relates to, among others, polynucleotides encoding the aforementioned fragments, polynucleotides that hybridize to polynucleotides encoding the fragments, particularly those that hybridize under stringent conditions, and polynucleotides, such as PCR primers, for amplifying polynucleotides that encode the fragments. In these regards, preferred polynucleotides are those that correspondent to the preferred fragments, as discussed above.

Vectors, host cells, expression

The present invention also relates to vectors which comprise a polynucleotide or polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

Host cells can be genetically engineered to incorporate polynucleotides and express polypeptides of the present invention. Introduction of a polynucleotide into the host cell can be affected by calcium phosphate transfection, DEAE-dextran mediated transfection, transfection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction, infection or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., *BASIC METHODS IN MOLECULAR BIOLOGY*, (1986) and Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

Polynucleotide constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

In accordance with this aspect of the invention the vector may be, for example, a plasmid vector, a single or double-stranded phage vector, a single or double-stranded RNA or DNA viral vector. Plasmids generally are designated herein by a lower case p preceded and/or followed by capital letters and/or numbers, in accordance with standard naming conventions that are familiar to those of skill in the art. Starting plasmids disclosed herein are either commercially available, publicly available, or can be constructed from available plasmids by routine application of well known, published procedures. Many plasmids and other cloning and expression vectors that can be used in accordance with the present invention are well known and readily available to those of skill in the art.

Preferred among vectors, in certain respects, are those for expression of polynucleotides and polypeptides of the present invention. Generally, such vectors comprise cis-acting control regions effective for expression in a host operatively linked to the polynucleotide to be expressed. Appropriate trans-acting factors either are supplied by the host, supplied by a complementing vector or supplied by the vector itself upon introduction into the host.

In certain preferred embodiments in this regard, the vectors provide for specific expression. Such specific expression may be inducible expression or expression only in certain types of cells or both inducible and cell-specific. Particularly preferred among inducible vectors are vectors that can be induced for expression by environmental factors that are easy to manipulate, such as temperature and nutrient additives. A variety of vectors suitable to this aspect of the invention, including constitutive and inducible expression vectors for use in prokaryotic and eukaryotic hosts, are well known and employed routinely by those of skill in the art.

A great variety of expression vectors can be used to express a polypeptide of the invention. Such vectors include, among others, chromosomal, episomal and virus-derived vectors, e.g., vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids, all may be used for expression in accordance with this aspect of the present invention. Generally, any vector suitable to maintain, propagate or express polynucleotides to express a polypeptide in a host may be used for expression in this regard.

The appropriate DNA sequence may be inserted into the vector by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

The DNA sequence in the expression vector is operatively linked to appropriate expression control sequence(s),

including, for instance, a promoter to direct mRNA transcription. Representatives of such promoters include, but are not limited to, the phage lambda PL promoter, the *E. coli* lac, trp and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs.

In general, expression constructs will contain sites for transcription initiation and termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs will include a translation initiating AUG at the beginning and a termination codon appropriately positioned at the end of the polypeptide to be translated.

In addition, the constructs may contain control regions that regulate as well as engender expression. Generally, in accordance with many commonly practiced procedures, such regions will operate by controlling transcription, such as transcription factors, repressor binding sites and termination, among others.

Vectors for propagation and expression generally will include selectable markers and amplification regions, such as, for example, those set forth in Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, *E. coli*, streptomyces and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

The following vectors, which are commercially available, are provided by way of example. Among vectors preferred for use in bacteria are pQE70, pQE60 and pQE9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16A, pNH18A, pNH46A, available from Stratagene; and ptc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia, and pBR322 (ATCC 37017). Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. These vectors are listed solely by way of illustration of the many commercially available and well known vectors that are available to those of skill in the art for use in accordance with this aspect of the present invention. It will be appreciated that any other plasmid or vector suitable for, for example, introduction, maintenance, propagation or expression of a polynucleotide or polypeptide of the invention in a host may be used in this aspect of the invention.

Promoter regions can be selected from any desired gene using vectors that contain a reporter transcription unit lacking a promoter region, such as a chloramphenicol acetyl transferase ("CAT") transcription unit, downstream of restriction site or sites for introducing a candidate promoter fragment; i.e., a fragment that may contain a promoter. As is well known, introduction into the vector of a promoter-containing fragment at the restriction site upstream of the cat gene engenders production of CAT activity, which can be detected by standard CAT assays. Vectors suitable to this end are well known and readily available, such as pKK232-8 and pCM7. Promoters for expression of polynucleotides of the present invention include not only well known and readily available promoters, but also promoters that readily may be obtained by the foregoing technique, using a reporter gene.

Among known prokaryotic promoters suitable for expression of polynucleotides and polypeptides in accordance with

the present invention are the *E. coli* lacI and lacZ and promoters, the T3 and T7 promoters, the gpt promoter, the lambda PR, PL promoters and the trp promoter.

Among known eukaryotic promoters suitable in this regard are the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous sarcoma virus ("RSV"), and metallothionein promoters, such as the mouse metallothionein-I promoter.

Recombinant expression vectors will include, for example, origins of replication, a promoter preferably derived from a highly-expressed gene to direct transcription of a downstream structural sequence, and a selectable marker to permit isolation of vector containing cells after exposure to the vector.

Polynucleotides of the invention, encoding the heterologous structural sequence of a polypeptide of the invention generally will be inserted into the vector using standard techniques so that it is operably linked to the promoter for expression. The polynucleotide will be positioned so that the transcription start site is located appropriately 5' to a ribosome binding site. The ribosome binding site will be 5' to the AUG that initiates translation of the polypeptide to be expressed. Generally, there will be no other open reading frames that begin with an initiation codon, usually AUG, and lie between the ribosome binding site and the initiation codon. Also, generally, there will be a translation stop codon at the end of the polypeptide and there will be a polyadenylation signal in constructs for use in eukaryotic hosts. Transcription termination signal appropriately disposed at the 3' end of the transcribed region may also be included in the polynucleotide construct.

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to the N- or C-terminus of the polypeptide to improve stability and persistence in the host cell, during purification or during subsequent handling and storage. Also, region also may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability or to facilitate purification, among others, are familiar and routine techniques in the art. A preferred fusion protein comprises a heterologous region from immunoglobulin that is useful to solubilize or purify polypeptides. For example, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another protein or part thereof. In drug discovery, for example, proteins have been fused with antibody Fc portions for the purpose of high-throughput screening assays to identify antagonists. See, D. Bennett et al., *Journal of Molecular Recognition*, 8: 52-58 (1995) and K. Johanson et al., *The Journal of Biological Chemistry*, 270, (16): 9459-9471 (1995).

Cells typically then are harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Mammalian expression vectors may comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation sites, splice donor and acceptor sites, transcriptional termination sequences, and 5' flanking non-transcribed sequences that are necessary for expression. In certain preferred embodiments in this regard DNA sequences derived from the SV40 splice sites, and the SV40 polyadenylation sites are used for required non-transcribed genetic elements of these types.

Topoisomerase III polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification. Well known techniques for refolding protein may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

Topoisomerase III polynucleotides and polypeptides may be used in accordance with the present invention for a variety of applications, particularly those that make use of the chemical and biological properties of Topoisomerase III. Additional applications relate to diagnosis and to treatment of disorders of cells, tissues and organisms. These aspects of the invention are illustrated further by the following discussion.

Polynucleotide assays

This invention is also related to the use of the Topoisomerase III polynucleotides to detect complementary polynucleotides such as, for example, as a diagnostic reagent. Detection of a bacterial Topoisomerase III in a eukaryote, particularly a mammal, and especially a human, will provide a diagnostic method that can add to, define or allow a diagnosis of a disease. Eukaryotes (herein also "individual(s)"), particularly mammals, and especially humans, infected by a Topoisomerase III producing bacterium may be detected at the DNA or RNA level by a variety of techniques. Nucleic acids for diagnosis may be obtained from an individual's cells and tissues, such as bone, blood, muscle, cartilage, and skin. Tissue biopsy and autopsy material is also preferred for samples from an individual to use in a diagnostic assay. The bacterial DNA may be used directly for detection or may be amplified enzymatically by using PCR prior to analysis. PCR (Saiki et al., *Nature* 324: 163-166 (1986)). RNA or cDNA may also be used in the same ways. As an example, PCR primers complementary to the nucleic acid encoding Topoisomerase III can be used to identify and analyze Topoisomerase III presence and expres-

sion. Using PCR, characterization of the strain of prokaryote present in a eukaryote, particularly a mammal, and especially a human, may be made by an analysis of the genotype of the prokaryote gene. For example, deletions and insertions can be detected by a change in size of the amplified product in comparison to the genotype of a reference sequence. Point mutations can be identified by hybridizing amplified DNA to radiolabeled Topoisomerase III RNA or alternatively, radiolabeled Topoisomerase III antisense DNA sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in melting temperatures.

Sequence differences between a reference gene and genes having mutations also may be revealed by direct DNA sequencing. In addition, cloned DNA segments may be employed as probes to detect specific DNA segments. The sensitivity of such methods can be greatly enhanced by appropriate use of PCR or another amplification method. For example, a sequencing primer is used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabeled nucleotide or by automatic sequencing procedures with fluorescent-tags.

Genetic typing of various strains of bacteria based on DNA sequence differences may be achieved by detection of alteration in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents. Small sequence deletions and insertions can be visualized by high resolution gel electrophoresis. DNA fragments of different sequences may be distinguished on denaturing formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial melting temperatures (see, e.g., Myers et al., *Science*, 230: 1242 (1985)).

Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (e.g., Cotton et al., *Proc. Nat'l. Acad. Sci., USA*, 85: 4397-4401 (1985)).

Thus, the detection of a specific DNA sequence may be achieved by methods such as hybridization, RNase protection, chemical cleavage, direct DNA sequencing or the use of restriction enzymes, (e.g., restriction fragment length polymorphisms ("RFLP") and Southern blotting of genomic DNA.

In addition to more conventional gel-electrophoresis and DNA sequencing, mutations also can be detected by in situ analysis.

Cells carrying mutations or polymorphisms in the gene of the present invention may also be detected at the DNA level by a variety of techniques, to allow for serotyping, for example. Nucleic acids for diagnosis may be obtained from an infected individual's cells, including but not limited to blood, urine, saliva, tissue biopsy and autopsy material or from bacteria isolated and cultured from the above sources. The bacterial DNA may be used directly for detection or may be amplified enzymatically by using PCR (Saiki et al., *Nature*, 324:163-166 (1986)) prior to analysis. RT-PCR can also be used to detect mutations. It is particularly preferred to used RT-PCR in conjunction with automated detection systems, such as, for example, GeneScan. RNA or cDNA may also be used for the same purpose, PCR or RT-PCR. As an example, PCR primers complementary to the nucleic acid encoding Topoisomerase III can be made using known methods and used to identify and analyze mutations. They may also be used to obtain full length gene sequence using known methods, such as, for example,

PCR. Examples of such primers include, but are not limited to, 5'-TAAAAGAACGTATGAGAAAG-3' [SEQ ID NO:4] (upper primer) and 5'-AAAAACAATACCAAAAGCGAACT-3' [SEQ ID NO:5] (lower primer). For example, deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabeled RNA or alternatively, radiolabeled antisense DNA sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in melting temperatures. These primers may be used for amplifying Topoisomerase III cDNA isolated from a sample derived from an individual. The invention also provides such primers with 1, 2, 3 or 4 nucleotides removed from the 5' and/or the 3' end. The primers may be used to amplify the gene isolated from the individual such that the gene may then be subject to various techniques for elucidation of the DNA sequence. In this way, mutations in the DNA sequence may be detected.

20 Polypeptide assays

The present invention also relates to a diagnostic assays such as quantitative and diagnostic assays for detecting levels of Topoisomerase III protein in cells and tissues, including determination of normal and abnormal levels.

Thus, for instance, a diagnostic assay in accordance with the invention for detecting expression of Topoisomerase III protein compared to normal control tissue samples may be used to detect the presence of an infection. Assay techniques that can be used to determine levels of a protein, such as an Topoisomerase III protein of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays. Among these ELISAs frequently are preferred. An ELISA assay initially comprises preparing an antibody specific to Topoisomerase III, preferably a monoclonal antibody. In addition a reporter antibody generally is prepared which binds to the monoclonal antibody. The reporter antibody is attached a detectable reagent such as radioactive, fluorescent or enzymatic reagent, in this example horseradish peroxidase enzyme.

Antibodies

The polypeptides, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. The present invention includes, for examples monoclonal and polyclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library.

Antibodies generated against the polypeptides corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide.

For preparation of monoclonal antibodies, any technique known in the art which provides antibodies produced by continuous cell line cultures can be used. Examples include various techniques, such as those in Kohler, G. and Milstein, C., *Nature* 256: 495-497 (1975); Kozbor et al., *Immunology Today* 4: 72 (1983); Cole et al., pg. 77-96 in *MONOCLONAL ANTIBODIES AND CANCER THERAPY*, Alan R. Liss, Inc. (1985).

Techniques described for the production of single chain antibodies (U.S. Pat. No. 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies to immunogenic polypeptide products of this invention.

Alternatively phage display technology could be utilized to select antibody genes with binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-Fbp or from naive libraries (McCafferty, J. et al., *Nature* 348, 552-554 (1990); Marks, J. et al., *Biotechnology* 10: 779-783 (1992). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. et al., *Nature* 352, 624-628 (1991).

If two antigen binding domains are present each domain may be directed against a different epitope—termed 'bispecific' antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or purify the polypeptide of the present invention by attachment of the antibody to a solid support for isolation and/or purification by affinity chromatography.

Thus among others, antibodies against Topoisomerase III may be employed to inhibit and/or treat infections, particularly bacterial infections, and especially Staphylococcal infections as well as to monitor the effectiveness of antibiotic treatment.

Polypeptide derivatives include antigenically, epitopically or immunologically equivalent derivatives which form a particular aspect of this invention. The term "antigenically equivalent derivative" as used herein encompasses a polypeptide or its equivalent which will be specifically recognized by certain antibodies which, when raised to the protein or polypeptide according to the present invention, interfere with the immediate physical interaction between pathogen and mammalian host. The term "immunologically equivalent derivative" as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to interfere with the immediate physical interaction between pathogen and mammalian host.

The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof is used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the polypeptide. The antigen may be associated, for example by conjugation, with an immunogenic carrier protein for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively a multiple antigenic peptide comprising multiple copies of the protein or polypeptide, or an antigenically or immunologically equivalent polypeptide thereof may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

Preferably the antibody or derivative thereof is modified to make it less immunogenic in the individual. For example, if the individual is human the antibody may most preferably be "humanised"; where the complementarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody, for example as described in Jones, P. et al., *Nature* 321: 522-525 (1986) or Tempest et al., *Biotechnology* 9: 266-273 (1991).

The use of a polynucleotide of the invention in genetic immunization will preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscles (Wolff et al., *Hum Mol Genet* 1:363 (1992), Man-

thorpe et al., *Hum. Gene Ther.* 1963:4: 419 (1963), delivery of DNA complexed with specific protein carriers (Wu et al., *J. Biol. Chem.* 264:16985 (1989), coprecipitation of DNA with calcium phosphate (Benvenisty and Reshef, *Proc. Nat'l. Acad. Sci. (USA)*, 83:9551 (1986), encapsulation of DNA in various forms of liposomes (Kaneda et al., *Science* 243:375 (1989), particle bombardment (Tang et al., *Nature*, (1992) 356:152, Eisenbraun et al., *DNA Cell Biol* 12:791 (1993) and in vivo infection using cloned retroviral vectors (Seeger et al., *Proc. Nat'l. Acad. Sci. (USA)* 81:5849 (1984). Topoisomerase III binding molecules and assays

This invention also provides a method for identification of molecules, such as binding molecules, that bind Topoisomerase III. Genes encoding proteins that bind Topoisomerase III, such as binding proteins, can be identified by numerous methods known to those of skill in the art, for example, ligand panning and FACS sorting. Such methods are described in many laboratory manuals such as, for instance, Coligan et al., *Current Protocols in Immunology* 1(2): Chapter 5 (1991).

For instance, expression cloning may be employed for this purpose. To this end polyadenylated RNA is prepared from a cell expressing Topoisomerase III, a cDNA library is created from this RNA, the library is divided into pools and the pools are transfected individually into cells that are not expressing Topoisomerase III. The transfected cells then are exposed to labeled Topoisomerase III. Topoisomerase III can be labeled by a variety of well-known techniques including standard methods of radio-iodination or inclusion of a recognition site for a site-specific protein kinase.) Following exposure, the cells are fixed and binding of Topoisomerase III is determined. These procedures conveniently are carried out on glass slides.

Pools are identified of cDNA that produced Topoisomerase III-binding cells. Sub-pools are prepared from these positives, transfected into host cells and screened as described above. Using an iterative sub-pooling and re-screening process, one or more single clones that encode the putative binding molecule, such as a binding molecule, can be isolated.

Alternatively a labeled ligand can be photoaffinity linked to a cell extract, such as a membrane or a membrane extract, prepared from cells that express a molecule that it binds, such as a binding molecule. Cross-linked material is resolved by polyacrylamide gel electrophoresis ("PAGE") and exposed to X-ray film. The labeled complex containing the ligand-binding can be excised, resolved into peptide fragments, and subjected to protein microsequencing. The amino acid sequence obtained from microsequencing can be used to design unique or degenerate oligonucleotide probes to screen cDNA libraries to identify genes encoding the putative binding molecule.

Polypeptides of the invention also can be used to assess Topoisomerase III binding capacity of Topoisomerase III binding molecules, such as binding molecules, in cells or in cell-free preparations.

Polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may be natural substrates and ligands or may be structural or functional mimetics.

Antagonists and Agonists—assays and molecules

As mentioned above, both increases and decreases in DNA density have been associated with bacterial responses to environmental challenges. Accordingly, modulating, i.e., agonizing or antagonizing, the appropriate response could result in a potential antibiotic effect.

The invention also provides a method of screening compounds to identify those which enhance or block the action of Topoisomerase III on cells, such as its interaction with substrate molecules, such as supercoiled DNA. Compounds which block the action of Topoisomerase III on cells include those which act as poisons and stabilize Topoisomerase III in a covalent complex with DNA, resulting in an inhibitory effect on cell growth. An antagonist is a compound which decreases the natural biological functions of Topoisomerase III. An agonist is a compound which increases the natural biological functions of Topoisomerase III.

Barrett et al., *Antimicrob. Agents Chemother.* 34:1 (1990) review in-vitro assays which can be used to measure inhibition of topoisomerases. These assays can be categorized as catalytic assays and noncatalytic assays. Catalytic assays for bacterial topoisomerase III include, for example, assays to measure the relaxation of supercoiled DNA. Noncatalytic assays, also known as 'cleavable complex' assays, measure the formation of a key covalent reaction intermediate. Froelich-Ammon and Osheroff *J. Biol. Chem.* 270:21429 (1995) review the mechanistic basis of noncatalytic assays of topoisomerase poisons.

Supercoiled DNA relaxation assay

To screen for inhibitors of the relaxation reaction, a candidate inhibitor and a preparation of Topoisomerase III are incubated with a supercoiled DNA substrate, for example plasmid or phage DNA, in an appropriate buffer containing Mg^{2+} , or an alternative divalent metal ion. Reaction products are separated by agarose gel electrophoresis, visualized by ethidium bromide staining, and quantified by densitometry.

DNA oligomer cleavage assay

A single stranded DNA oligomer containing appropriate cleavage sites, for example the 22mer GAATGAGCCG-CAACTTCGGGAT (SEQ ID NO: 3), or an appropriately labelled derivative, may be used as substrate. An appropriate label may be a radiolabel or a fluorescent chromophore attached at the 5' or 3' end of the oligo, according to the specific assay used. The substrate is incubated with a candidate inhibitor and a preparation of Topoisomerase III, in an appropriate buffer. The buffer may contain Mg^{2+} or an alternative divalent metal ion. Mg^{2+} is not essential for the cleavage reaction, although its inclusion may be desirable to facilitate the interaction of certain classes of inhibitors. The reaction is stopped by the addition of an appropriate denaturant, for example 1% SDS or 100 mM NaOH. Generation of the cleavable complex (stabilization of the key covalent reaction intermediate) may be measured by a number of methods. For example, electrophoresis using a denaturing polyacrylamide gel can be used to separate the 5' labelled cleaved DNA product which may then be quantified by densitometry. Alternatively, the 3' labelled DNA product may be assayed by virtue of its covalent association with Topoisomerase III. This may be performed by the SDS/K precipitation assay, in which radiolabelled DNA associated with precipitated protein is measured, or by a capture assay format in which Topoisomerase III is immobilized using an antibody and the amount of associated labelled DNA is measured.

Whole cell assays

Topoisomerase III-like effects of potential agonists and antagonists and poisons, may be measured, for instance, by determining activity of a reporter system that is sensitive to alterations in gene expression following interaction of the candidate molecule with a cell or appropriate cell preparation. Reporter systems that may be useful in this regard include but are not limited to colorimetric labeled substrate

converted into product, a reporter gene that is responsive to changes in Topoisomerase III activity, and binding assays known in the art.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polypeptide of the invention and thereby inhibit or extinguish its activity, or stabilize the key covalent reaction intermediate with DNA. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a binding molecule, without inducing Topoisomerase III-induced activities, thereby preventing the action of Topoisomerase III by excluding Topoisomerase III from binding.

Potential antagonists include a small molecule which binds to and occupies the binding site of the polypeptide thereby preventing binding to cellular binding molecules, such as binding molecules, such that normal biological activity is prevented. Examples of small molecules include but are not limited to small organic molecules, peptides or peptide-like molecules.

Other potential antagonists include antisense molecules. Antisense technology can be used to control gene expression through antisense DNA or RNA or through double- or triple-helix formation. Antisense techniques are discussed, for example, in—Okano, *J. Neurochem* 56: 560 (1991); *OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION*, CRC Press, Boca Raton, Fla. (1988). Triple helix formation is discussed in, for instance Lee et al., *Nucleic Acids Research* 6: 3073 (1979); Cooney et al., *Science* 241: 456 (1988); and Dervan et al., *Science* 251: 1360 (1991). The methods are based on binding of a polynucleotide to a complementary DNA or RNA. For example, the 5' coding portion of a polynucleotide that encodes the mature polypeptide of the present invention may be used to design an antisense RNA oligonucleotide of from about 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription thereby preventing transcription and the production of Topoisomerase III. The antisense RNA oligonucleotide hybridizes to the mRNA in vivo and blocks translation of the mRNA molecule into Topoisomerase III polypeptide. The oligonucleotides described above can also be delivered to cells such that the antisense RNA or DNA may be expressed in vivo to inhibit production of Topoisomerase III.

Preferred potential antagonists include compounds related to and derivatives of each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein upon expression can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding the amino terminal regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest.

The antagonists and agonists of the invention may be employed in a composition with a pharmaceutically acceptable carrier, e.g., as hereinafter described.

The antagonists and agonists may be employed for instance to inhibit staphylococcal infections. Vaccines

Another aspect of the invention relates to a method for inducing an immunological response in an individual, particularly a mammal which comprises inoculating the individual with Topoisomerase III, or a antigenic fragment or variant thereof, adequate to produce antibody to protect said

individual from infection, particularly bacterial infection and most particularly Staphylococcal infection. Yet another aspect of the invention relates to a method of inducing immunological response in an individual which comprises, through gene therapy, delivering gene encoding Topoisomerase III, or an antigenic fragment or a variant thereof, for expressing Topoisomerase III, or a fragment or a variant thereof in vivo in order to induce an immunological response to produce antibody to protect said individual from disease.

A further aspect of the invention relates to an immunological composition which, when introduced into a host capable or having induced within it an immunological response, induces an immunological response in such host to a Topoisomerase III or protein coded therefrom, wherein the composition comprises a recombinant Topoisomerase III or protein coded therefrom comprising DNA which codes for and expresses an antigen of said Topoisomerase III or protein coded therefrom.

The Topoisomerase III or a fragment thereof may be fused with co-protein which may not by itself produce antibodies, but is capable of stabilizing the first protein and producing a fused protein which will have immunogenic and protective properties. Thus fused recombinant protein, preferably further comprises an antigenic co-protein, such as Glutathione-S-transferase (GST) or beta-galactosidase, relatively large co-proteins which solubilise the protein and facilitate production and purification thereof. Moreover, the co-protein may act as an adjuvant in the sense of providing a generalized stimulation of the immune system. The co-protein may be attached to either the amino or carboxy terminus of the first protein.

The present invention also includes a vaccine formulation which comprises the immunogenic recombinant protein together with a suitable carrier. Since the protein may be broken down in the stomach, it is preferably administered parenterally, including, for example, administration that is subcutaneous, intramuscular, intravenous, or intradermal. Formulations suitable for parenteral administration include aqueous and non-aqueous sterile injection solutions which may contain anti-oxidants, buffers, bacteriostats and solutes which render the formulation isotonic with the bodily fluid, preferably the blood, of the individual; and aqueous and non-aqueous sterile suspensions which may include suspending agents or thickening agents. The formulations may be presented in unit-dose or multi-dose containers, for example, sealed ampoules and vials and may be stored in a freeze-dried condition requiring only the addition of the sterile liquid carrier immediately prior to use. The vaccine formulation may also include adjuvant systems for enhancing the immunogenicity of the formulation, such as oil-in water systems and other systems known in the art. The dosage will depend on the specific activity of the vaccine and can be readily determined by routine experimentation.

Whilst the invention has been described with reference to certain Topoisomerase III, it is to be understood that this covers fragments of the naturally occurring protein and similar proteins (for example, having sequence homologies of 50% or greater) with additions, deletions or substitutions which do not substantially affect the immunogenic properties of the recombinant protein.

Compositions

The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or the agonists or antagonists. Thus, the polypeptides of the present invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues

or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol and combinations thereof. The formulation should suit the mode of administration.

Kits

The invention further relates to diagnostic and pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, reflecting approval by the agency of the manufacture, use or sale of the product for human administration.

Administration

Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

The pharmaceutical compositions generally are administered in an amount effective for treatment or prophylaxis of a specific indication or indications. In general, the compositions are administered in an amount of at least about 10 $\mu\text{g/kg}$ body weight. In most cases they will be administered in an amount not in excess of about 8 mg/kg body weight per day. Preferably, in most cases, dose is from about 10 $\mu\text{g/kg}$ to about 1 mg/kg body weight, daily. It will be appreciated that optimum dosage will be determined by standard methods for each treatment modality and indication, taking into account the indication, its severity, route of administration, complicating conditions and the like.

In therapy or as a prophylactic, the active agent may be administered to an individual as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

For administration to mammals, and particularly humans, it is expected that the daily dosage level of the active agent will be from 0.01 mg/kg to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual and will vary with the age, weight and response of the particular individual. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

In-dwelling devices include surgical implants, prosthetic devices and catheters, i.e., devices that are introduced to the

body of an individual and remain in position for an extended time. Such devices include, for example, artificial joints, heart valves, pacemakers, vascular grafts, vascular catheters, cerebrospinal fluid shunts, urinary catheters, continuous ambulatory peritoneal dialysis (CAPD) catheters, etc.

The composition of the invention may be administered by injection to achieve a systemic effect against relevant bacteria shortly before insertion of an in-dwelling device. Treatment may be continued after surgery during the in-body time of the device. In addition, the composition could also be used to broaden perioperative cover for any surgical technique to prevent Staphylococcal wound infections.

Many orthopedic surgeons consider that humans with prosthetic joints should be considered for antibiotic prophylaxis before dental treatment that could produce a bacteraemia. Late deep infection is a serious complication sometimes leading to loss of the prosthetic joint and is accompanied by significant morbidity and mortality. It may therefore be possible to extend the use of the active agent as a replacement for prophylactic antibiotics in this situation.

In addition to the therapy described above, the compositions of this invention may be used generally as a wound treatment agent to prevent adhesion of bacteria to matrix proteins exposed in wound tissue and for prophylactic use in dental treatment as an alternative to, or in conjunction with, antibiotic prophylaxis.

Alternatively, the composition of the invention may be used to bathe an indwelling device immediately before insertion. The active agent will preferably be present at a concentration of 1 µg/ml to 10 mg/ml for bathing of wounds or indwelling devices.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response.

A suitable unit dose for vaccination is 0.5–5 microgram/kg of antigen, and such dose is preferably administered 1–3 times and with an interval of 1–3 weeks.

With the indicated dose range, no adverse toxicological effects will be observed with the compounds of the invention which would preclude their administration to suitable individuals.

The antibodies described above may also be used as diagnostic reagents to detect the presence of bacteria containing Topoisomerase.

Each reference disclosed herein is incorporated by reference herein in its entirety. Any patent application to which this application claims priority is also incorporated by reference herein in its entirety.

In order to facilitate understanding of the following example certain frequently occurring methods and/or terms will be described.

EXAMPLES

The present invention is further described by the following examples. The examples are provided solely to illustrate the invention by reference to specific embodiments. These exemplification's, while illustrating certain specific aspects of the invention, do not portray the limitations or circumscribe the scope of the disclosed invention.

Certain terms used herein are explained in the foregoing glossary.

All examples were carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. Routine molecular biology techniques of the following examples can be carried out as described in standard laboratory manuals, such

as Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

All parts or amounts set out in the following examples are by weight, unless otherwise specified.

Unless otherwise stated size separation of fragments in the examples below was carried out using standard techniques of agarose and polyacrylamide gel electrophoresis ("PAGE") in Sambrook et al., *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989) and numerous other references such as, for instance, by Goeddel et al., *Nucleic Acids Res.* 8: 4057 (1980).

Unless described otherwise, ligations were accomplished using standard buffers, incubation temperatures and times, approximately equimolar amounts of the DNA fragments to be ligated and approximately 10 units of T4 DNA ligase ("ligase") per 0.5 microgram of DNA.

The polynucleotide having the DNA sequence given in (SEQ ID NO: 1) was obtained from the sequencing of a library of clones of chromosomal DNA of *Staphylococcus aureus* WCUH 29 in *E. coli*.

To obtain the polynucleotide encoding the Topoisomerase III protein using the DNA sequence given in (SEQ ID NO: 1) typically a library of clones of chromosomal DNA of *Staphylococcus aureus* WCUH 29 in *E. coli* or some other suitable host is probed with a radiolabelled oligonucleotide, preferably a 17 mer or longer, derived from the partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using high stringency washes. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E. F. and Sambrook et al., *MOLECULAR CLONING, A LABORATORY MANUAL*, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). (see Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70).

Example 1

Isolation of DNA coding for Novel Topoisomerase III Protein from *Staphylococcus aureus*

The polynucleotide having the DNA sequence given in (SEQ ID NO: 1) was obtained from a library of clones of chromosomal DNA of *Staphylococcus aureus* in *E. coli*. In some cases the sequencing data from two or more clones containing overlapping *Staphylococcus aureus* DNA was used to construct the contiguous DNA sequence in (SEQ ID NO: 1). Libraries may be prepared Libraries may be prepared by routine methods, for example, Methods 1 and 2 below.

Total cellular DNA is isolated from *Staphylococcus aureus* strain WCUH 29 according to standard procedures and size-fractionated by either of two methods.

Method 1

Total cellular DNA is mechanically sheared by passage through a needle in order to size-fractionate according to standard procedures. DNA fragments of up to 11 kbp in size are rendered blunt by treatment with exonuclease and DNA polymerase, and EcoRI linkers added. Fragments are ligated into the vector Lambda ZapII that has been cut with EcoRI, the library packaged by standard procedures and *E. coli* infected with the packaged library. The library is amplified by standard procedures.

Method 2

Total cellular DNA is partially hydrolysed with a combination of four restriction enzymes (RsaI, PstI, AluI and Bsh1235) and size-fractionated according to standard procedures. EcoRI linkers are ligated to the DNA and the fragments then ligated into the vector Lambda ZapII that have been cut with EcoRI, the library packaged by standard procedures, and *E. coli* infected with the packaged library. The library is amplified by standard procedures.

Example 2

Characterization of Topoisomerase III Gene Expression

a) Isolation of *Staphylococcus aureus* WCUH29 RNA from infected tissue samples

Infected tissue samples, in 2-ml cyro-storage tubes, are removed from -80°C . storage into a dry ice ethanol bath. In a microbiological safety cabinet the samples are disrupted up to eight at a time while the remaining samples are kept frozen in the dry ice ethanol bath. To disrupt the bacteria within the tissue sample, 50–100 mg of the tissue is transferred to a FastRNA tube containing a silica/ceramic matrix (BIO101). Immediately, 1 ml of extraction reagents (FastRNA reagents, BIO101) are added to give a sample to reagent volume ratio of approximately 1 to 20. The tubes are shaken in a reciprocating shaker (FastPrep FP20, BIO101) at 6000 rpm for 20–120 sec. The crude RNA preparation is extracted with chloroformisoamyl alcohol, and precipitated with DEPC-treated/Isopropanol Precipitation Solution (BIO101). RNA preparations are stored in this isopropanol solution at -80°C . if necessary. The RNA is pelleted (12,000 g for 10 min.), washed with 75% ethanol (v/v in DEPC-treated water), air-dried for 5–10 min, and resuspended in 0.1 ml of DEPC-treated water.

Quality of the RNA isolated is assessed by running samples on 1% agarose gels. 1× TBE gels stained with ethidium bromide are used to visualise total RNA yields. To demonstrate the isolation of bacterial RNA from the infected tissue 1× MOPS, 2.2M formaldehyde gels are run and vacuum blotted to Hybond-N (Amersham). The blot is then hybridized with a 32 P labelled oligonucleotide probe specific to 16S rRNA of *Staphylococcus aureus* (K. Greisen, M. Loeffelholz, A. Purohit and D. Leong, J. Clin. (1994) Microbiol. 32 335–351). An oligonucleotide of the sequence: 5'-gctcctaaagggtactccaccggc-3' is used as a probe. The size of the hybridizing band is compared to that of control RNA isolated from in vitro grown *Staphylococcus aureus* WCUH29 in the Northern blot. Correct sized bacterial 16S rRNA bands can be detected in total RNA samples which show extensive degradation of the mammalian RNA when visualised on TBE gels.

b) The removal of DNA from *Staphylococcus aureus* WCUH29-derived RNA

DNA was removed from 50 ng samples of RNA by a 30 minute treatment at 37°C . with 10 units of RNAase-free DNAaseI (GeneHunter) in the buffer supplied in a final volume of 57 microliters.

The DNAase was inactivated and removed by phenol:chloroform extraction. RNA was precipitated with 5 microliters of 3 M NaOAc and 200 microliters 100% EtOH, and pelleted by centrifugation at 12,000 g for 10 minutes. The RNA is pelleted (12,000 g for 10 min.), washed with 75% ethanol (v/v in DEPC-treated water), air-dried for 5–10 min, and resuspended in 10–20 microliters of DEPC-treated water. RNA yield is quantitated by OD₂₆₀ after 1:1000 dilution of the cleaned RNA sample. RNA is stored at -80°C . if necessary and reverse-transcribed within one week. c) The preparation of cDNA from RNA samples derived from infected tissue

10 microliter samples of DNAase treated RNA are reverse transcribed using a SuperScript Preamplification System for First Strand cDNA Synthesis kit (Gibco BRL, Life Technologies) according to the manufacturers instructions. 1 nanogram of random hexamers is used to prime each reaction. Controls without the addition of SuperScriptII reverse transcriptase are also run. Both +/-RT samples are treated with RNaseH before proceeding to the PCR reaction.

d) The use of PCR and fluorogenic probes to determine the presence of a bacterial cDNA species

Specific sequence detection occurs by amplification of target sequences in the PE Applied Biosystems 7700 Sequence Detection System in the presence of an oligonucleotide probe labeled at the 5' and 3' ends with a reporter and quencher fluorescent dye, respectively (FQ probe), which anneals between the two PCR primers. Only specific product will be detected when the probe is bound between the primers. As PCR amplification proceeds, the 5'-nuclease activity of Taq polymerase initially cleaves the reporter dye from the probe. The signal generated when the reporter dye is physically separated from the quencher dye is detected by measuring the signal with an attached CCD camera. Each signal generated equals one probe cleaved which corresponds to amplification of one target strand PCR reactions are set up using the PE Applied Biosystem TaqMan PCR Core Reagent Kit according to the instructions supplied such that each reaction contains 5 microliters 10× PCR Buffer II, 7 microliters 25 mM MgCl₂, 5 microliters 300 nM forward primer, 5 microliters reverse primer, 5 microliters specific FQ probe, 1 microliter each 10 mM dATP, 10 mM dCTP, 10 mM dGTP and 20 mM dUTP, 13.25 microliters distilled water, 0.5 microliters AmpErase UNG, and 0.25 microliters AmpliTaq DNA polymerase to give a total volume of 45 microliters.

Amplification proceeds under the following thermal cycling conditions: 50°C . hold for 2 minutes, 95°C . hold for 10 minutes, 40 cycles of 95°C . for 15 seconds and 60°C . for 1 minute, followed by a 25°C . hold until sample is retrieved. Detection occurs real-time. Data is collected at the end of the reaction.

RT/PCR controls may include +/-reverse transcriptase reactions, amplification along side genes known to be transcribed under the conditions of study and amplification of 1 microgram of genomic DNA.

Primer pairs and corresponding probes which fail to generate signal in DNA PCR or RT/PCR are PCR failures and as such are uninformative. Of those which generate signal with DNA PCR, two classes are distinguished in RT/PCR: 1. Genes which are not transcribed in vivo reproducibly fail to generate signal in RT/PCR; and 2. Genes which are transcribed in vivo reproducibly generate signal in RT/PCR and show a stronger signal in the +RT samples than the signal (if at all present) in -RT controls. Based on these analyses it was discovered that *S. aureus* topoisomerase III gene was expressed in vivo.

Primers used for Example 2 are as follows:

topB fwd primer GTTATACGATATGATTGTCGAGCGT [SEQ ID NO:6]
 topB rev primer GTGCCCTGCAACCTCTAAACT [SEQ ID NO:7]
 topB probe FAM-CCTCCGCACGAGTATGACGCG-TAMRA [SEQ ID NO:8]

FAM and TAMRA labeling of primers and the uses of 10 21:3761-3766; Livak, K J, Flood, S J A, Marmaro, J., Giusti, W, and Deetz, K. 1995. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods and Applications 4:357-362.).

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 8

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2803 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

TTCATTGTAC TGGTGAGGAA GTTTATATAA GTATGATGCT GATTTCATAAT TCGAATGTTC    60
AATGAACGAT TTTATTGTG TAATATCATA TAACTGAACT ATGCTCATGT CATTACCTCC    120
GTACTTTTGT TTAATTTTAT TATATAGTAT TTCAACTGAA ATGAAAGTTA ATAGTGATAT    180
TAACATGTTA CAATACATTT AACACCATTG AATTTAAATC AAAGATTAGT GGAATAGATG    240
AAGCACGTTT GAAATAAAAG AACGTATGAG AAAGGATAAT TTATGAAATC TTTAATATTA    300
GCTGAAAAC CATCAGTTGC AAGAGATATT GCTGATGCTT TACAATAAAT TCAGAAGCGT    360
AATGGTTACT TTGAAAATAA CCAATATATT GTCACGTGGG CGTTAGGTCA TCTAGTGACA    420
AATGCGACAC CTGAACAATA CGATAAAAT TTAAGGAAT GCGATTAGA AGACCTTCCA    480
ATTATACCTA AATATATGAA AACTGTTGTT ATTGGTAAAA CAAGCAAACA ATTTAAAACA    540
GTAAAAGCGT TAATTTTAGA TAATAAAGTG AAAGATATTA TTATTGCAAC AGATGCTGGA    600
CGAGAAGGTG AACTAGTTGC AAGATTGATT TTGGATAAAG TTGGTAACAA AAAGCCAATC    660
CGTCGATTAT GGATTAGCTC AGTTACTAAA AAAGCTATTC AACAAGGTTT TAAAAATTTA    720
AAAGACGGTC GTCAATATAA CGATTGTGAT TATGCAGCGT TAGCGAGAAG CGAGGCAGAT    780
TGGATTGTTG GGATTAAATG AACGCGTGCA CTAACAACAA AGTATGATGC ACAGCTATCC    840
CTGGGACGTG TTCAGACACC AACGATTCAA TTAGTAAATA CACGACAACA AGAGATTAAT    900
CAGTTCAAAC CACAACAATA CTTTACATTA TCATTAACGG TAAAAGGTTT TGATTTTCAG    960
CTAGAATCAA ATCAGCGATA TACCAATAAA GAACTTTAG AACAGATGGT TAATAATTTG    1020
AAAAATGTCG ATGGTAAGAT TAAATCTGTT GCTACTAAAC ATAAGAAGTC GTATCCGCAA    1080
TCACTGTACA ATTTAACAGA TTTACAACAA GATATGTATA GACGTTATAA AATTGGACCT    1140
AAAGAAACAT TGAATACACT TCAAAGCTTA TATGAGAGAC ATAAAGTCGT AACCTATCCA    1200

```

-continued

```

AGAACAGATT CAAACTATTT AACAACTGAT ATGGTAGATA CTATGAAAGA ACGTATTTCAG 1260
GCGACGATGG CAACAACATA TAAAGACCAA GCACGCCCAT TAATGTCTAA AACATTTTCA 1320
TCAAAAATGT CGATATTTAA TAATCAAAAA GTATCTGATC ACCATGCAAT TATTCCTACA 1380
GAAGTGAGAC CTGTCATGTC AGACTTAAGT AATAGAGAAT TAAAGTTATA CGATATGATT 1440
GTCGAGCGTT TTTTAGAAGC TTTAATGCCT CCGCACGAGT ATGACGCGAT AACTGTAAC 1500
TTAGAGTTG CAGGGCACAC ATTTGTTTG AAAGAGAATG TAACAACTGT TTTAGGTTTT 1560
AAATCTATTA GACAAGGTGA ATCTATTACA GAGATGCAAC AGCCTTTTTC AGAAGGCGAT 1620
GAAGTGAAGA TTTCAAAAAC AAACATTAGA GAACATGAAA CAACACCTCC AGAATATTTT 1680
AATGAAGGTT CGTTATTAAA AGCGATGGAG AACCCCTCAGA ACTTTATTCA ATTGAAGGAT 1740
AAAAAATATG CGCAAACTTT AAAACAAACA GGTGGTATCG GCACAGTTGC AACCAAGGCC 1800
GACATTATCG ATAAATTAT TAATATGAAT GCCATTGAAT CAAGAGACGG TAAATTTAAA 1860
GTAACGTCAA AAGGTAACA AATATTAGAA TTAGCACCAG AAGAATTAAC GTCGCCACTT 1920
TTAACTGCAC AATGGGAAGA AAAATTACTT TTAATTGAAC GTGGTAAATA TCAGGCGAAA 1980
ACATTTATTA ATGAAATGAA AGATTTTACG AAAGATGTTG TAAATGGGAT TAAAAATAGT 2040
GATCGTAAAT ATAAACACGA TAATTTAACA ACCACAGAAT GCCCAACGTG TGGTAAATTC 2100
ATGATTAAAG TTAAACTAA AAATGGTCAG ATGCTTGTGT GCCAAGATCC ATCTTGTAG 2160
ACGAAAAAGA ATGTACAGCG CAAAACAAAT GCAAGATGTC CAAACTGTAA AAAGAAATTA 2220
ACGTTGTTTG GTAAAGGGAA AGAAGCGGTA TATCGTTGTG TTTGTGGACA TTCTGAAACG 2280
CAAGCAGATA TGGATCAGCG TATGAAGTCT AAATCCTCTG GTAAAGTATC TCGTAAAGAA 2340
ATGAAAAAGT ATATGAATAA AAATGAAGGT TTAGACAATA ATCCGTTTAA AGATGCATTA 2400
AAGAAGTTGA ATTTATAGAT AAAATCGAAC AAAGTTGAAT CAGAAAAACG AAAAGTTTCG 2460
TTTTGGTATT GTTTTTTATT AAGAATGATA TTAACTATT AAGGTATTTT AAAAAAGGA 2520
GCATCCATTC GTGAAAAACT ATTTCCAGTT CGATAAATAT GGAACAACT TTTAAAGAGA 2580
AATCTTAGGC GGTATCACAA CTTTCTTATC TATGGCCTAT ATTTTAGCAG TTAACCCGCA 2640
AGTTTAAAGT TTAGCAGGTG TTAAGGCGT ATCAGAAGAT ATGAAAATGG ACCAAGGTGC 2700
CATTTTGTGA GCGACTGCAT TAGCAGCATT TGTAGGCTCG CTATTCATGG GACTAATAGC 2760
TAAATATCCA ATCGCATTAG CACCAGGTAT GGGATTGGAA TTC 2803

```

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 711 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

Met Lys Ser Leu Ile Leu Ala Glu Lys Pro Ser Val Ala Arg Asp Ile
 1             5             10             15
Ala Asp Ala Leu Gln Ile Asn Gln Lys Arg Asn Gly Tyr Phe Glu Asn
 20             25             30
Asn Gln Tyr Ile Val Thr Trp Ala Leu Gly His Leu Val Thr Asn Ala
 35             40             45
Thr Pro Glu Gln Tyr Asp Lys Asn Leu Lys Glu Trp Arg Leu Glu Asp
 50             55             60

```

-continued

Leu	Pro	Ile	Ile	Pro	Lys	Tyr	Met	Lys	Thr	Val	Val	Ile	Gly	Lys	Thr	65	70	75	80
Ser	Lys	Gln	Phe	Lys	Thr	Val	Lys	Ala	Leu	Ile	Leu	Asp	Asn	Lys	Val	85	90	95	
Lys	Asp	Ile	Ile	Ile	Ala	Thr	Asp	Ala	Gly	Arg	Glu	Gly	Glu	Leu	Val	100	105	110	
Ala	Arg	Leu	Ile	Leu	Asp	Lys	Val	Gly	Asn	Lys	Lys	Pro	Ile	Arg	Arg	115	120	125	
Leu	Trp	Ile	Ser	Ser	Val	Thr	Lys	Lys	Ala	Ile	Gln	Gln	Gly	Phe	Lys	130	135	140	
Asn	Leu	Lys	Asp	Gly	Arg	Gln	Tyr	Asn	Asp	Leu	Tyr	Tyr	Ala	Ala	Leu	145	150	155	160
Ala	Arg	Ser	Glu	Ala	Asp	Trp	Ile	Val	Gly	Ile	Asn	Ala	Thr	Arg	Ala	165	170	175	
Leu	Thr	Thr	Lys	Tyr	Asp	Ala	Gln	Leu	Ser	Leu	Gly	Arg	Val	Gln	Thr	180	185	190	
Pro	Thr	Ile	Gln	Leu	Val	Asn	Thr	Arg	Gln	Gln	Glu	Ile	Asn	Gln	Phe	195	200	205	
Lys	Pro	Gln	Gln	Tyr	Phe	Thr	Leu	Ser	Leu	Thr	Val	Lys	Gly	Phe	Asp	210	215	220	
Phe	Gln	Leu	Glu	Ser	Asn	Gln	Arg	Tyr	Thr	Asn	Lys	Glu	Thr	Leu	Glu	225	230	235	240
Gln	Met	Val	Asn	Asn	Leu	Lys	Asn	Val	Asp	Gly	Lys	Ile	Lys	Ser	Val	245	250	255	
Ala	Thr	Lys	His	Lys	Lys	Ser	Tyr	Pro	Gln	Ser	Leu	Tyr	Asn	Leu	Thr	260	265	270	
Asp	Leu	Gln	Gln	Asp	Met	Tyr	Arg	Arg	Tyr	Lys	Ile	Gly	Pro	Lys	Glu	275	280	285	
Thr	Leu	Asn	Thr	Leu	Gln	Ser	Leu	Tyr	Glu	Arg	His	Lys	Val	Val	Thr	290	295	300	
Tyr	Pro	Arg	Thr	Asp	Ser	Asn	Tyr	Leu	Thr	Thr	Asp	Met	Val	Asp	Thr	305	310	315	320
Met	Lys	Glu	Arg	Ile	Gln	Ala	Thr	Met	Ala	Thr	Thr	Tyr	Lys	Asp	Gln	325	330	335	
Ala	Arg	Pro	Leu	Met	Ser	Lys	Thr	Phe	Ser	Ser	Lys	Met	Ser	Ile	Phe	340	345	350	
Asn	Asn	Gln	Lys	Val	Ser	Asp	His	His	Ala	Ile	Ile	Pro	Thr	Glu	Val	355	360	365	
Arg	Pro	Val	Met	Ser	Asp	Leu	Ser	Asn	Arg	Glu	Leu	Lys	Leu	Tyr	Asp	370	375	380	
Met	Ile	Val	Glu	Arg	Phe	Leu	Glu	Ala	Leu	Met	Pro	Pro	His	Glu	Tyr	385	390	395	400
Asp	Ala	Ile	Thr	Val	Thr	Leu	Glu	Val	Ala	Gly	His	Thr	Phe	Val	Leu	405	410	415	
Lys	Glu	Asn	Val	Thr	Thr	Val	Leu	Gly	Phe	Lys	Ser	Ile	Arg	Gln	Gly	420	425	430	
Glu	Ser	Ile	Thr	Glu	Met	Gln	Gln	Pro	Phe	Ser	Glu	Gly	Asp	Glu	Val	435	440	445	
Lys	Ile	Ser	Lys	Thr	Asn	Ile	Arg	Glu	His	Glu	Thr	Thr	Pro	Pro	Glu	450	455	460	
Tyr	Phe	Asn	Glu	Gly	Ser	Leu	Leu	Lys	Ala	Met	Glu	Asn	Pro	Gln	Asn	465	470	475	480
Phe	Ile	Gln	Leu	Lys	Asp	Lys	Lys	Tyr	Ala	Gln	Thr	Leu	Lys	Gln	Thr				

-continued

485	490	495
Gly Gly Ile Gly Thr Val Ala Thr Arg Ala Asp Ile Ile Asp Lys Leu 500	505	510
Phe Asn Met Asn Ala Ile Glu Ser Arg Asp Gly Lys Ile Lys Val Thr 515	520	525
Ser Lys Gly Lys Gln Ile Leu Glu Leu Ala Pro Glu Glu Leu Thr Ser 530	535	540
Pro Leu Leu Thr Ala Gln Trp Glu Glu Lys Leu Leu Leu Ile Glu Arg 545	550	555
Gly Lys Tyr Gln Ala Lys Thr Phe Ile Asn Glu Met Lys Asp Phe Thr 565	570	575
Lys Asp Val Val Asn Gly Ile Lys Asn Ser Asp Arg Lys Tyr Lys His 580	585	590
Asp Asn Leu Thr Thr Thr Glu Cys Pro Thr Cys Gly Lys Phe Met Ile 595	600	605
Lys Val Lys Thr Lys Asn Gly Gln Met Leu Val Cys Gln Asp Pro Ser 610	615	620
Cys Lys Thr Lys Lys Asn Val Gln Arg Lys Thr Asn Ala Arg Cys Pro 625	630	635
Asn Cys Lys Lys Lys Leu Thr Leu Phe Gly Lys Gly Lys Glu Ala Val 645	650	655
Tyr Arg Cys Val Cys Gly His Ser Glu Thr Gln Ala His Met Asp Gln 660	665	670
Arg Met Lys Ser Lys Ser Ser Gly Lys Val Ser Arg Lys Glu Met Lys 675	680	685
Lys Tyr Met Asn Lys Asn Glu Gly Leu Asp Asn Asn Pro Phe Lys Asp 690	695	700
Ala Leu Lys Asn Leu Asn Leu 705	710	

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 22 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAATGAGCCG CAACTTCGGG AT

22

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TAAAGAACG TATGAGAAAG

20

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 23 base pairs

-continued

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AAAAACAATA CCAAAAGCGA ACT 23

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GTTATACGAT ATGATTGTCG AGCGT 25

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GTGCCCTGCA ACCTCTAAAG T 21

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CCTCCGCACG AGTATGACGC G 21

What is claimed is:

1. An isolated polynucleotide segment comprising: a first polynucleotide sequence, wherein the first polynucleotide sequence (a) is a reference sequence that encodes the amino acid sequence set forth in SEQ ID NO:2, or (b) has at least 95% identity with the reference sequence, wherein said identity is determined using an algorithm selected from the group consisting of BLASTP, BLASTN or FASTA, where the algorithm is adapted to give the largest match between sequences tested, over the entire length of the reference sequence.

2. An isolated polynucleotide segment according to claim 1 comprising: a first polynucleotide sequence, wherein the first polynucleotide sequence (a) encodes a reference sequence that has the amino acid sequence set forth in SEQ ID NO:2, or (b) encodes a polypeptide sequence identical with the reference sequence except that, over the entire length corresponding to the reference sequence, the encoded polypeptide sequence has up to ten amino acid substitutions, amino acid deletions or amino acid insertions.

3. The isolated polynucleotide segment of claim 1, comprising the first polynucleotide sequence wherein the first polynucleotide sequence is (a) identical with the reference sequence, or (b) has at least 97% identity with the reference sequence, wherein said identity is determined using an algorithm selected from the group consisting of BLASTP, BLASTN or FASTA, where the algorithm is adapted to give the largest match between sequences tested, over the entire length of the reference sequence.

4. The isolated polynucleotide segment of claim 1, wherein the first polynucleotide sequence encodes a topoisomerase III polypeptide.

5. An isolated polynucleotide segment comprising the full complement of the entire length of the first polynucleotide sequence of claim 1.

6. The isolated polynucleotide segment of claim 5, wherein the first polynucleotide sequence (a) encodes a reference sequence that has the amino acid sequence set forth in SEQ ID NO:2, or (b) encodes a polypeptide

sequence identical with the reference sequence except that, over the entire length corresponding to the reference sequence, the encoded polypeptide sequence has up to five amino acid substitutions, amino acid deletions or amino acid insertions.

7. The isolated polynucleotide segment of claim 6, wherein the first polynucleotide sequence encodes a topoisomerase III polypeptide.

8. A vector comprising the polynucleotide segment of claim 1.

9. A vector comprising the polynucleotide segment of claim 5.

10. An isolated host cell transformed with the polynucleotide segment of claim 1 to express the first polynucleotide sequence.

11. A process for producing an topoisomerase III polypeptide of the first polynucleotide sequence comprising the step of culturing the host cell of claim 10 under conditions sufficient for the production of said polypeptide, which is encoded by the first polynucleotide sequence.

12. An isolated polynucleotide segment comprising a first polynucleotide sequence, wherein the first polynucleotide sequence is (a) a first reference sequence that encodes the amino acid sequence set forth in SEQ ID NO:2, or (b) has at least 90% identity with the first reference sequence, wherein said identity is determined using an algorithm selected from the group consisting of BLASTP, BLASTN or FASTA, where the algorithm is adapted to give the largest match between sequences tested, over the entire length of the first reference sequence

wherein the first polynucleotide sequence is (a) a second reference sequence which encodes the same mature polypeptide, expressed by the topoisomerase III gene contained in *Staphylococcus aureus* WCUH 29 contained in NCIMB Deposit No. 40771, or (b) has at least 95% identity with the second reference sequence, wherein said identity is determined using an algorithm selected from the group consisting of BLASTP, BLASTN or FASTA, where the algorithm is adapted to give the largest match between sequences tested, over the entire length of the second reference sequence.

13. An isolated polynucleotide segment of claim 12, wherein the first polynucleotide sequence is (a) the second reference sequence, or (b) has at least 97% identity with the second reference sequence, wherein said identity is determined using an algorithm selected from the group consisting of BLASTP, BLASTN or FASTA, where the algorithm is adapted to give the largest match between sequences tested, over the entire length of the second reference sequence.

14. An isolated polynucleotide segment of claim 13 comprising a first polynucleotide sequence encoding the same mature polypeptide expressed by the topoisomerase III gene contained in *Staphylococcus aureus* WCUH 29 contained in NCIMB Deposit No. 40771, or the full complement of the entire length of such first polynucleotide sequence.

15. A polynucleotide encoding a fusion polypeptide including a polynucleotide segment according to claim 12.

16. The isolated polynucleotide segment of claim 1, wherein the first polynucleotide sequence is identical with a third reference sequence which (a) is nucleotides 283 to 2415 inclusive of the polynucleotide sequence set forth in SEQ ID NO:1, or (b) has at least 95% identity with the third reference sequence, wherein said identity is determined using an algorithm selected from the group consisting of BLASTP, BLASTN or FASTA, where the algorithm is adapted to give the largest match between sequences tested, over the entire length of the third reference sequence;

wherein the first polynucleotide sequence hybridizes under stringent conditions to a polynucleotide sequence which is the full complement of said third reference sequence, wherein stringent conditions means hybridization will occur only if there is at least 95% identity between the polynucleotide sequences to be hybridized.

17. An isolated polynucleotide segment of claim 16, wherein the first polynucleotide sequence (a) is the sequence from nucleotides 283 to 2415 inclusive of the polynucleotide sequence set forth in SEQ ID NO:1, or (b) has at least 97% identity with the third reference sequence, wherein said identity is determined using an algorithm selected from the group consisting of BLASTP, BLASTN or FASTA, where the algorithm is adapted to give the largest match between sequences tested, over the entire length of the third reference sequence.

18. The isolated polynucleotide segment of claim 1, wherein the first polynucleotide sequence hybridizes under stringent conditions to a polynucleotide sequence which is the full complement of a third reference polynucleotide having nucleotides 283 to 2415 of SEQ ID NO:1.

19. A recombinant polynucleotide segment comprising nucleotides 283 to 2415 of the polynucleotide sequence set forth in SEQ ID NO:1, or the full complement of the entire length of the nucleotide sequence set forth in SEQ ID NO:1.

20. A recombinant polynucleotide segment which encodes a polypeptide comprising a region having the amino acid sequence of SEQ ID NO:2.

21. A vector comprising the recombinant polynucleotide segment of claim 20.

22. An isolated host cell transformed with the recombinant polynucleotide segment of claim 20 to express the recombinant polynucleotide segment.

23. A process for producing a topoisomerase III polypeptide of the polynucleotide sequence comprising the step of culturing a host cell of claim 22 under conditions sufficient for the production of said polypeptide.

24. The isolated polynucleotide of claim 1, wherein said isolated polynucleotide encodes a topoisomerase III polypeptide that is involved in altering DNA topology in a bacterial cell.

25. The isolated polynucleotide of claim 16, wherein said isolated polynucleotide encodes a topoisomerase III polypeptide that is involved in altering DNA topology in a bacterial cell.

* * * * *



US006645499B1

(12) **United States Patent**
Lal et al.

(10) **Patent No.:** **US 6,645,499 B1**
 (45) **Date of Patent:** ***Nov. 11, 2003**

(54) **HUMAN ENA/VASP-LIKE PROTEIN SPLICE VARIANT**

(75) **Inventors:** **Preeti Lal**, Santa Clara, CA (US); **Karl J. Guegler**, Menlo Park, CA (US); **Neil C. Corley**, Mountain View, CA (US)

(73) **Assignee:** **Incyte Corporation**, Palo Alto, CA (US)

(*) **Notice:** Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

This patent is subject to a terminal disclaimer.

(21) **Appl. No.:** **09/387,811**

(22) **Filed:** **Sep. 1, 1999**

Related U.S. Application Data

(62) Division of application No. 09/227,420, filed on Jan. 8, 1999, now Pat. No. 5,990,087, which is a division of application No. 09/026,587, filed on Feb. 20, 1998, now Pat. No. 5,912,128.

(51) **Int. Cl.⁷** **A61K 39/00**

(52) **U.S. Cl.** **424/185.1; 530/350**

(58) **Field of Search** **514/12; 530/350; 424/185.1**

(56) **References Cited**
PUBLICATIONS

Lerner, RA Tapping the immunological repertoire to produce antibodies of predetermined specificity. *Nature* 299:592-596, Oct. 1982.*
 Gertler, F.B. et al., "Mena, a Relative of VASP and Drosophila Enabled, Is Implicated in the Control of Microfilament Dynamics" *Cell* (1996) 87:227-239.
 Ermekova, K.S. et al., "The WW Domain of Neural Protein FE65 Interacts with Proline-rich Motifs in Mena, the Mammalian Homolog of Drosophila Enabled" *J.Biol.Chem.* (1997) 272:32869-32877.

Linial, M., "Proline clustering in proteins from synaptic vesicles" *Neuroreport* (1994) 5:2009-2015.

Ohta, S. et al., "Differential Display Cloning of a Novel Rat cDNA (RNB6) That Shows High Expression in the Neonatal Brain Revealed a Member of Ena/VASP Family" *Biochem. Biophys. Res. Commun.* (1997) 237:307-312.

Gertler, F.B. et al., (GI 1644452), GenBank Sequence Database (Accession U72519), National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland, 20894 (1996).

Walter, U., (GI 624963), GenBank Sequence Database (Accession Z46389), National Center for Biotechnology Information, National Library of Medicine, Bethesda, Maryland, 20894. (1995)

Haffner, C. et al., "Molecular cloning, structural analysis and functional expression of the proline-rich focal adhesion and microfilament-associated protein VASP" *EMBO J.* (1995) 14:19-27.

Adams, M.D., et al., GenBank Accession No. AA312190, cross reference to *Nature* 377, supp. 28:3-17, 1995.

Naeve, C.W., et al., Accuracy of Automated DNA sequencing: a multi-laboratory comparison of sequencing results. *Biotechniques*. 19(3) :448-453. (1995).

Gertler, F.B., et al., GenBank Accession No. U72519, cross referenced to *Cell* 87(2):227-239 (1996).

Bork, *Genome Research*, vol. 10: 398-400, 2000.*

Attwood, *Science* vol. 290: 471-473, 2000.*

Skolnich et al. *TIBTECH* vol. 18: pp. 34-39, 2000.*

* cited by examiner

Primary Examiner—Patrick J. Nolan

(74) *Attorney, Agent, or Firm*—Incyte Corporation

(57) **ABSTRACT**

The invention provides a human ena/VASP-like protein splice variant (EVL1) and polynucleotides which identify and encode EVL1. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for treating or preventing disorders associated with expression of EVL1.

2 Claims, 10 Drawing Sheets

```

5' TTT AAG TAG GCT ATA AAA ATC AAG TTG CTG TCT TCA GAG GGT CTG TGG TCC TCT 54
    9      18      27      36      45
    63 CAT AGG CTG GTG GGA GTA CAG GAC TCG CCT CAG GGT TCC CTG TGC 108
    72      81      90      99
    117 TTT TCA GCC ATG GCC ACA AGT GAA CAG AGT ATC TGC CAA GCC CGG GCT 162
    126      135      144      153
    171 ATG GTC TAC GAT GAC ACC AGT AAG AAA TGG GTA CCA ATC AAA CCT GGC 216
    180      189      198      207
    225 GGA TTC AGC CGG ATC AAC ATC TAC CAC AAC ACT GCC AGC AAC ACC TTC 270
    234      243      252      261
    279 GTC GTT GGA GTC AAG TTG CAG GAT CAG CAG GTT GTG ATC AAT TAT TCA ATC 324
    288      297      306      315
    333 GGG CTG AAG TAC AAT CAG GCC ACG CCA ACC TTC CAC CAG TGG CGA GAT 378
    342      351      360      369

```

FIGURE 1A

387	GCC CGC CAG GTC TAC GGC TTA AAC TTT GCA AGT AAA GAA GAG GCA ACC ACC TTC	405	414	423	432
A R Q	V Y G L N F A S K E A T T F				
441	TCC AAT GCA ATG CTG TTT GCC CTG AAC ATC ATG AAT TCC CAA GAA GGC CCC	459	468	477	486
S N A	M L F A L N I M N S Q E G P				
495	TCC AGC CAG CGT CAG GTG CAG AAT GGC CCC TCT CCT GAT GAG ATG GAC ATC CAG	513	522	531	540
S S Q	R Q V Q N G P S P D E M D I Q				
549	AGA AGA CAA GTG ATG GAG CAG CAC CAG CAG CGT CAG GAA TCT CTA GAA AGA	567	576	585	594
R R Q	V M E Q H Q Q Q R Q E S L E R				
603	AGA ACC TCG GCC ACA GGG CCC ATC CTC CCA CCA GGA CAT CCT TCA TCT GCA GCC	621	630	639	648
R T S	A T G P I L P P G H P S S A A				
657	AGC GCC CCC GTC TCA TGT AGT GGG CCT CCA CCG CCC CCC CCT CTA GTC CCA	675	684	693	702
S A P	V S C S S G P P P P P P P L V P				
711	CCT CCA CCC ACT GGG GCT ACC CCA CCT CCC CCA CCC CCA CTG CCA GCC GGA GGA	729	738	747	756
P P P	T G A T P P P P P P P L P A G G				

FIGURE 1B

765 GCC CAG GGG TCC AGC CAC GAC GAG AGC TCC ATG TCA GGA CTG GCC GCT GCC ATA
A Q G S S H D E S S M S G L A A A I 810

819 GCT GGG GCC AAG CTG AGA AGA GTC CAA CGG CCA GAA GAC GCA TCT GGA GGC TCC
A G A K L R R V Q R P E D A S G G S 864

873 AGT CCC AGT GGG ACC TCA AAG TCC GAT GCC AAC CGG GCA AGC AGC GGC GGT GGC
S P S G T S S K S D A N R A S S G G 918

927 GGA GGA GGC CTC ATG GAG GAA ATG AAC AAA CTG CTG GCC AAG AGG AGA AAA GCA
G G G L M E E M N K L L A K R R K A 972

981 GCC TCC CAG TCA GAC AAG CCA GCC GAG AAG AAG GAA GAT GAA AGC CAA ATG GAA
A S Q S D K P A E K K E D E S Q M E 1026

1035 GAT CCT AGT ACC TCC CCC TCT CCG GGG ACC CGA GCA GCC AGC CAG CCA CCT AAC
D P S T S P S P G T R A A S Q P P N 1080

1089 TCC TCA GAG GCT GGC CGG AAG CCC TGG GAG CGG AGC AAC TCG GTG GAG AAG CCT
S S E A G R K P W E R S N S V E K P 1134

FIGURE 1C

1143	1152	1161	1170	1179	1188
GTG TCC TCG ATT CTG TCC AGA ACC CCG TCT GTG GCA AAG AGC CCC GAA GCT AAG					
V S S I L S R T P S V A K S P E A K					
1197	1206	1215	1224	1233	1242
AGC CCC CTT CAG TCG CAG CCT CAC TCT AGG ATG AAG CCT GCT GGG AGC GTG AAT					
S P L Q S Q P H S R M K P A G S V N					
1251	1260	1269	1278	1287	1296
GAC ATG GCC CTG GAT GCC TTC GAC TTG GAC CGG ATG AAG CAG GAG ATC CTA GAG					
D M A L D A F D L D R M K Q E I L E					
1305	1314	1323	1332	1341	1350
GAG GTG GTG AGA GAG CTC CAC AAG GTG AAG GAG GAG ATC ATC GAC GCC ATC AGG					
E V V R E L H K V K E E I I D A I R					
1359	1368	1377	1386	1395	1404
CAG GAG CTG AGT GGG ATC AGC ACC ACG TAA GGG GCC GGC CTC GCT GCG CTG ATT					
Q E L S G I S T T					
1413	1422	1431	1440	1449	1458
CGT CGA GCC CAT CCG GCG ACA GAG GAC AGC CAG AAG CCC AGC CAG CCC CAG ACT					
1467	1476	1485	1494	1503	1512
CCA GTG CAC CAG AGC ACG CAC AGG AGC CTG GCG GCG CTG CTG TGA AAC GTC CTG					

FIGURE 1D

1521	1530	1539	1548	1557	1566
ACC TGT GAT CAC ACA TGA CAG TGA GGA AAC CAA GTG CAA CTC CTG GGT TTT TTT					
1575	1584	1593	1602	1611	1620
TAG ATT CTG CCT GAC ACG GAA CAC CAG GTC TGC TCG TCT TTT TTG TGT TTT ATA					
1629	1638	1647	1656	1665	1674
TTT GCT TAT TTA AGG TAC ATT TCT TTG GGT TTC TAG AGA CGC CCC TAA GTC ACC					
1683	1692	1701	1710	1719	1728
TGC TTC ATT AGA CGG TTT CCA GGT TTT CTC CCA GGT GAC GCT GTT AGC GCC TCA					
1737	1746	1755	1764	1773	1782
GCT GGC GGT GAC AGC CGG CCC AGC GTG GCG CCA CCA CAC ACC GCA GAG CTG TCC					
1791	1800	1809	1818	1827	1836
AGG CAC AGC TCC GTC CCC AGC GCT CAT GGT GGT GAA ACT GTC TGT CAT GCA CCA					
1845	1854	1863	1872	1881	
CGG TGT CTG TGT CCA CAC AGT AAT AAA CGG TTT ACT GTC CGC AAA AAA AA 3					

FIGURE 1E

1	M	A	T	S	E	Q	S	I	C	Q	A	R	A	S	V	M	V	Y	D	D	T	S	K	K	W	V	P	I	K	P	3089412
1	M	-	-	S	E	Q	S	I	C	Q	A	R	A	S	V	M	V	Y	D	D	T	S	K	K	W	V	P	I	K	P	GI 1644453
1	M	-	-	S	E	T	V	I	C	S	S	R	A	T	V	M	L	Y	D	D	G	N	K	R	W	L	P	A	G	T	GI 624964
31	G	Q	Q	G	F	S	R	I	N	I	Y	H	N	T	A	S	N	T	F	R	V	V	G	V	K	L	Q	-	D	Q	3089412
29	G	Q	Q	G	F	S	R	I	N	I	Y	H	N	T	A	S	S	T	F	R	V	V	G	V	K	L	Q	-	D	Q	GI 1644453
29	G	P	Q	A	F	S	R	V	Q	I	Y	H	N	P	T	A	N	S	F	R	V	V	G	R	K	M	Q	P	D	Q	GI 624964
60	Q	V	V	I	N	Y	S	I	V	K	G	L	K	Y	N	Q	A	T	P	T	F	H	Q	W	R	D	A	R	Q	V	3089412
58	Q	V	V	I	N	Y	S	I	V	K	G	L	K	Y	N	Q	A	T	P	T	F	H	Q	W	R	D	A	R	Q	V	GI 1644453
59	Q	V	V	I	N	C	A	I	V	R	G	V	K	Y	N	Q	A	T	P	N	F	H	Q	W	R	D	A	R	Q	V	GI 624964
90	Y	G	L	N	F	A	S	K	E	E	A	T	T	F	S	N	A	M	L	F	A	L	N	I	M	N	S	Q	E	G	3089412
88	Y	G	L	N	F	A	S	K	E	E	A	T	T	F	S	N	A	M	L	F	A	L	N	I	M	N	S	Q	E	G	GI 1644453
89	W	G	L	N	F	G	S	K	E	D	A	A	Q	F	A	A	G	M	A	S	A	L	E	A	L	E	G	G	P	GI 624964	
120	G	P	S	S	Q	-	-	-	R	Q	V	Q	N	G	P	S	P	D	E	M	D	I	Q	R	R	Q	V	M	E	Q	3089412
118	G	P	S	T	Q	-	-	-	R	Q	V	Q	N	G	P	S	P	E	E	M	D	I	Q	R	R	Q	V	M	E	Q	GI 1644453
119	P	P	P	A	L	P	T	W	S	V	P	N	G	P	S	P	E	E	V	E	Q	Q	K	R	Q	-	-	-	-	GI 624964	
147	H	Q	Q	Q	R	Q	E	S	L	E	R	R	T	S	A	T	G	P	I	L	P	P	G	H	P	S	S	A	A	S	3089412
145	-	-	Q	H	R	Q	E	S	L	E	R	R	I	S	A	T	G	P	I	L	P	P	G	H	P	S	S	A	A	S	GI 1644453
145	-	Q	P	G	P	S	E	H	I	E	R	R	V	S	N	A	G	G	-	-	P	P	A	P	P	A	G	G	P	P	GI 624964

FIGURE 2A

349	A	K	S	P	E	A	K	S	P	L	Q	S	Q	P	H	S	R	M	K	P	A	G	S	V	N	D	M	A	L	D	3089412
341	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	K	P	A	G	S	V	N	D	V	G	L	D	GI 1644453
321	-	K	S	S	S	V	T	S	E	T	Q	P	C	T	P	S	S	-	-	-	-	-	-	-	-	-	-	-	-	D	GI 624964
379	A	F	D	L	D	R	M	K	Q	E	I	L	E	E	V	V	R	E	L	H	K	V	K	E	E	I	I	D	A	I	3089412
354	A	L	D	L	D	R	M	K	Q	E	I	L	E	E	V	V	R	E	L	H	K	V	K	E	E	I	I	D	A	I	GI 1644453
341	Y	S	D	L	Q	R	V	K	Q	E	L	L	E	E	V	K	K	E	L	Q	K	V	K	E	E	I	I	E	A	F	GI 624964
409	R	Q	E	L	S	G	I	S	T	T																					3089412
384	R	Q	E	L	S	G	I	S	T	T																					GI 1644453
371	V	Q	E	L	R	K	R	G	S	P																					GI 624964

FIGURE 2C

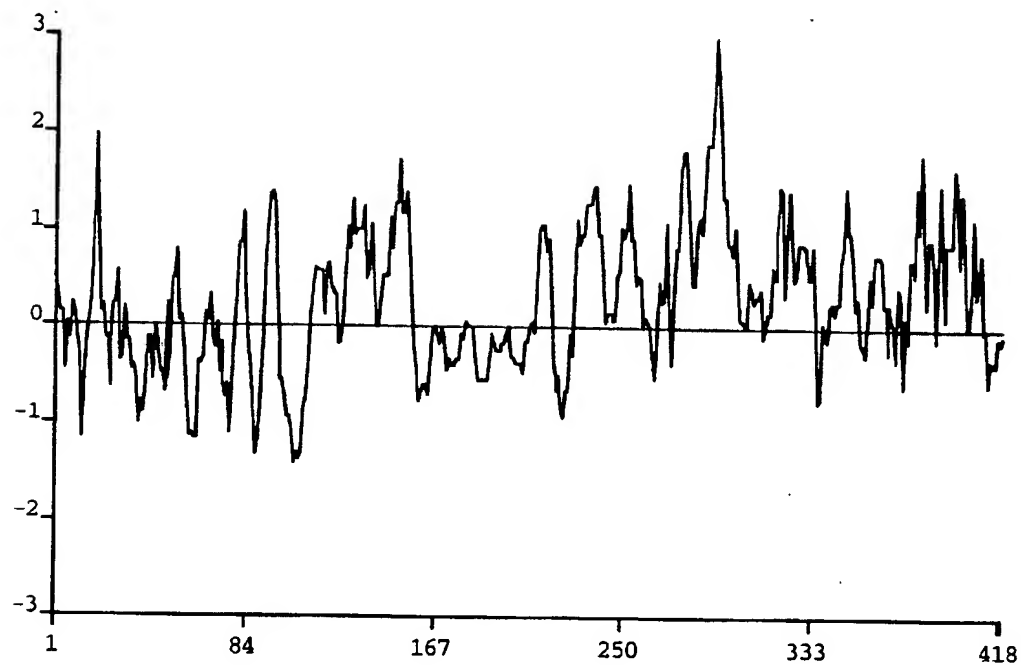


FIGURE 3A

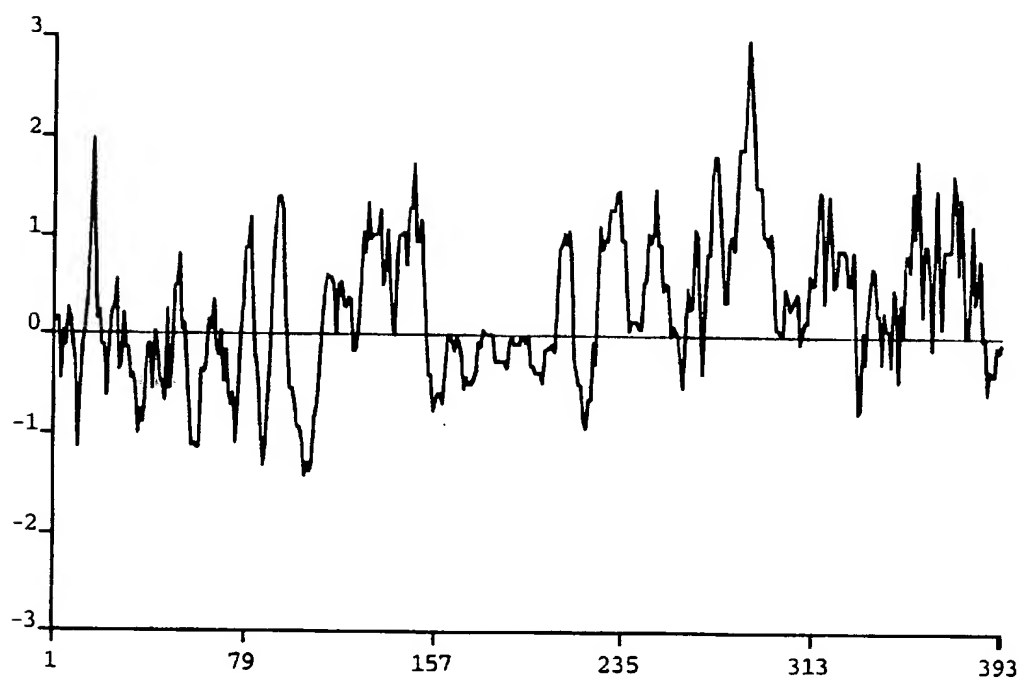


FIGURE 3B

HUMAN ENA/VASP-LIKE PROTEIN SPLICE VARIANT

This application is a divisional application of U.S. application Ser. No. 09/227,420, filed Jan. 8, 1999, issued as U.S. Pat. No. 5,990,087 which is a divisional of U.S. application Ser. No. 09/026,587; filed Feb. 20, 1998, issued Jun. 15, 1999, as U.S. Pat. No. 5,912,128.

FIELD OF THE INVENTION

This invention relates to nucleic acid and amino acid sequences of a human ena/VASP-like protein splice variant and to the use of these sequences in the diagnosis, treatment, and prevention of reproductive, immunological, vesicle trafficking, nervous system, developmental, and neoplastic disorders.

BACKGROUND OF THE INVENTION

The control of cell morphology and motility requires the coupling of external stimuli to processes that alter the cytoskeletal architecture. The mechanical forces that drive morphological change and migration arise initially from the microfilament-based cytoskeleton. A large body of evidence links various signal transduction pathways to the formation of cellular outgrowths. The migration of neuronal growth cones is a well-studied mechanism for the actin-driven formation of membrane protrusions. In one example, the processes of axonal outgrowth are mediated by the *Drosophila* homolog of the c-Abl tyrosine kinase (Abl) and the product of the Disabled gene (Dab). Homozygous mutants of Abl and Dab make few or no proper axonal connections. The defects caused by loss of Abl and Dab in *Drosophila* are ameliorated by mutations in the Enabled (Ena) gene. Ena protein is tyrosine phosphorylated and has a proline-rich core which binds to the SH3 domains of Abl protein and Src protein *in vitro*. The murine homolog of Ena (Mena) and ena/VASP-like protein have recently been described and are members of a family of related molecules that include vasoactive-stimulated phosphoprotein (VASP). These proteins share three distinct regions of similarity: an amino-terminal 115 amino acids (EVH1 domain); a proline-rich core; and a carboxy-terminal 226 amino acids (EVH2 domain). Mena has phosphotyrosine and phosphoserine moieties and binds Abl and Src SH3 domains. (Gertler, F. B. et al. (1996) *Cell* 87:227-239.)

Human platelet activation is inhibited by agents such as prostaglandins and nitric oxide donors, which elevate intracellular cAMP or cGMP levels. Activation of platelets is associated with increased formation of intracellular F-actin. VASP is an abundant *in vivo* substrate for cyclic nucleotide-dependant protein kinases in platelets. VASP is a ligand for profilin, an actin-monomer binding protein that can stimulate the formation of F-actin. VASP is organized into three distinct domains. A central proline-rich domain contains a GPPPPP motif as a single copy and as a 3-fold tandem repeat, as well as three conserved phosphorylation sites for cyclic nucleotide-dependent protein kinases. A C-terminal domain contains a repetitive mixed-charge cluster which is predicted to form an alpha-helix. VASP expression in transiently transfected BHK21 cells was predominantly detected at stress fibers, at focal adhesions, and in F-actin-containing cell surface protrusions. In contrast, truncated VASP lacking the C-terminal domain was no longer concentrated at focal adhesions. These data indicate that the C-terminal domain is required for anchoring VASP at focal adhesion sites, while the central domain may mediate VASP interaction with

profilin. (Ermeikova, K. S. et al. (1997) *J. Biol. Chem.* 272:32869-32877.)

In comparison, Mena binds FE65, a neuronal protein which binds to the cytoplasmic portion of the β -amyloid precursor protein (β -APP). β -APP is a precursor to β -amyloid peptide, the major constituent of the extracellular plaques present in brain tissue from Alzheimer disease patients. Both VASP and Mena bind their respective adapter proteins (profilin or FE65) via distinct proline-rich regions and thus regulate adapter interaction(s) with other molecules. (Ermeikova, K. S. et al, *supra*.) Proline-rich domains and proline clusters have been identified in many proteins, particularly those that are associated with synaptic vesicles and other secretory organelles. These domains and clusters act as protein-protein interaction modules. (Linial, M. (1994) *Neuroreport* 5:2009-2015.) Two members of ena/VASP-like proteins have been recently isolated from mouse and rat and share 98.5% sequence identity. (Gertler, *supra*; and Ohta, S. et al. (1997) *Biochem. Biophys. Res. Comm.* 237:307-312.)

The discovery of a new human ena/VASP-like protein splice variant and the polynucleotides encoding it satisfies a need in the art by providing new compositions which are useful in the diagnosis, treatment, and prevention of reproductive, immunological, vesicle trafficking, nervous system, developmental, and neoplastic disorders.

SUMMARY OF THE INVENTION

The invention is based on the discovery of a new human ena/VASP-like protein splice variant (EVL1), the polynucleotides encoding EVL1, and the use of these compositions for the diagnosis, treatment, or prevention of reproductive, immunological, vesicle trafficking, nervous system, developmental, and neoplastic disorders.

The invention features a substantially purified polypeptide, comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention further provides a substantially purified variant having at least 90% amino acid identity to the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. The invention also provides an isolated and purified polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. The invention also includes an isolated and purified polynucleotide variant having at least 90% polynucleotide identity to the polynucleotide encoding the polypeptide consisting of the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

Additionally, the invention provides a composition comprising a polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. The invention further provides an isolated and purified polynucleotide which hybridizes under stringent conditions to the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, as well as an isolated and purified polynucleotide which is complementary to the polynucleotide encoding the polypeptide consisting of the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides an isolated and purified polynucleotide comprising a sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2, and an isolated and purified polynucleotide variant having at least 90% polynucleotide identity to the polynucleotide comprising the sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2. The invention also provides an isolated and purified polynucleotide which

is complementary to the polynucleotide comprising the sequence of SEQ ID NO:2 or a fragment of SEQ ID NO:2.

The invention further provides an expression vector containing at least a fragment of the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1. In another aspect, the expression vector is contained within a host cell.

The invention also provides a method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, the method comprising the steps of: (a) culturing the host cell containing an expression vector containing at least a fragment of a polynucleotide encoding the polypeptide under conditions suitable for the expression of the polypeptide; and (b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition consisting of a substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1 in conjunction with a suitable pharmaceutical carrier.

The invention further includes a purified antibody which binds to a polypeptide consisting of the sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1, as well as a purified agonist and a purified antagonist of the polypeptide.

The invention also provides a method for treating or preventing a reproductive disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing an immunological disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing a vesicle trafficking disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing a nervous system disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing a developmental disorder, the method comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition comprising substantially purified polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for treating or preventing a neoplastic disorder, the method comprising administering to a subject in need of such treatment an effective amount of an antagonist of the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1.

The invention also provides a method for detecting a polynucleotide encoding a polypeptide comprising the

amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1 in a biological sample containing nucleic acids, the method comprising the steps of: (a) hybridizing the complement of the polynucleotide encoding the polypeptide comprising the amino acid sequence of SEQ ID NO:1 or a fragment of SEQ ID NO:1 to at least one of the nucleic acids of the biological sample, thereby forming a hybridization complex; and (b) detecting the hybridization complex, wherein the presence of the hybridization complex correlates with the presence of a polynucleotide encoding the polypeptide in the biological sample. In one aspect, the nucleic acids of the biological sample are amplified by the polymerase chain reaction prior to the hybridizing step.

BRIEF DESCRIPTION OF THE FIGS.

FIGS. 1A, 1B, 1C, 1D, and 1E show the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:2) of EVL1. The alignment was produced using MACDMASIS PRO software (Hitachi Software Engineering Co. Ltd., San Bruno, Calif.).

FIGS. 2A, 2B, and 2C show the amino acid sequence alignments among EVL1 (3089412; SEQ ID NO:1), mouse cna/VASP-like protein (GI 1644453; SEQ ID NO:3), and human VASP (GI 624964; SEQ ID NO:4), produced using the multi sequence alignment program of DNASTAR software (DNASTAR Inc., Madison Wiss.).

FIGS. 3A and 3B show the hydrophobicity plots for EVL1 (SEQ ID NO:1) and mouse cna/VASP-like protein (SEQ ID NO:3), respectively; the positive X axis reflects amino acid position and the negative Y axis reflects hydrophobicity (MACDMASIS PRO software).

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"EVL1," as used herein, refers to the amino acid sequences of substantially purified EVL1 obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and preferably the human species, from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist," as used herein, refers to a molecule which, when bound to EVL1, increases or prolongs the duration of the effect of EVL1. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of EVL1.

An "allele" or an "allelic sequence," as these terms are used herein, is an alternative form of the gene encoding EVL1. Alleles may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding EVL1, as described herein, include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide the same EVL1 or a polypeptide with at least one functional characteristic of EVL1. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding EVL1, and improper or unexpected hybridization to alleles, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding EVL1. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent EVL1. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of EVL1 is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine.

The terms "amino acid" or "amino acid sequence," as used herein, refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. In this context, "fragments", "immunogenic fragments", or "antigenic fragments" refer to fragments of EVL1 which are preferably about 5 to about 15 amino acids in length and which retain some biological activity or immunological activity of EVL1. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification," as used herein, relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction

(PCR) technologies well known in the art. (See, e.g., Dieffenbach, C. W. and G. S. Dveksler (1995) *PCR Primer, a Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y., pp.1-5.)

The term "antagonist," as it is used herein, refers to a molecule which, when bound to EVL1, decreases the amount or the duration of the effect of the biological or immunological activity of EVL1. Antagonists may include proteins, nucleic acids, carbohydrates, antibodies, or any other molecules which decrease the effect of EVL1.

As used herein, the term "antibody" refers to intact molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding the epitopic determinant. Antibodies that bind EVL1 polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant," as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (given regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense," as used herein, refers to any composition containing a nucleic acid sequence which is complementary to a specific nucleic acid sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form duplexes and to block either transcription or translation. The designation "negative" can refer to the antisense strand, and the designation "positive" can refer to the sense strand.

As used herein, the term "biologically active," refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic EVL1, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity," as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base pairing. For example, the sequence "A—G—T" binds to the complementary sequence "T—C—A." Complementarity between two single-stranded molecules may be "partial," such that only some of the nucleic acids bind, or it may be "complete," such that total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of the hybridization between the nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding

between nucleic acids strands, and in the design and use of peptide nucleic acid (PNA) molecules.

A "composition comprising a given polynucleotide sequence" or a "composition consisting of a given amino acid sequence," as these terms are used herein, refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation, an aqueous solution, or a sterile composition. Compositions comprising polynucleotide sequences encoding EVL1 or fragments of EVL1 may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

The phrase "consensus sequence," as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, extended using XL-PCR (Perkin Elmer, Norwalk, Conn.) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly, such as the GELVIEW Fragment Assembly system (GCG, Madison, Wis.). Some sequences have been both extended and assembled to produce the consensus sequence.

As used herein, the term "correlates with expression of a polynucleotide" indicates that the detection of the presence of nucleic acids, the same or related to a nucleic acid sequence encoding EVL1, by northern analysis is indicative of the presence of nucleic acids encoding EVL1 in a sample, and thereby correlates with expression of the transcript from the polynucleotide encoding EVL1.

A "deletion," as the term is used herein, refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative," as used herein, refers to the chemical modification of EVL1, of a polynucleotide sequence encoding EVL1, or of a polynucleotide sequence complementary to a polynucleotide sequence encoding EVL1. Chemical modifications of a polynucleotide sequence can include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

The term "homology," as used herein, refers to a degree of complementarity. There may be partial homology or complete homology. The word "identity" may substitute for the word "homology." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization, and the like) under conditions of reduced stringency. A substantially homologous sequence or hybridization probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of reduced stringency. This is not to say that conditions of reduced stringency are such that non-specific binding is permitted, as reduced stringency conditions require that the binding of

two sequences to one another be a specific (i.e., a selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% homology or identity). In the absence of non-specific binding, the substantially homologous sequence or probe will not hybridize to the second non-complementary target sequence.

The phrases "percent identity" or "% identity" refer to the percentage of sequence similarity found in a comparison of two or more amino acid or nucleic acid sequences. Percent identity can be determined electronically, e.g., by using the MEGALIGN program (Lasergene software package, DNASTAR, Inc., Madison, Wis.). The MEGALIGN program can create alignments between two or more sequences according to different methods, e.g., the clustal Method. (Higgins, D. G. and P. M. Sharp (1988) *Gene* 73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. The percentage similarity between two amino acid sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no homology between the two amino acid sequences are not included in determining percentage similarity. Percent identity between nucleic acid sequences can also be calculated by the clustal method, or by other methods known in the art, such as the Jotun Hein Method. (See, e.g., Hein, J. (1990) *Methods in Enzymology* 183:626-645.) Identity between sequences can also be determined by other methods known in the art, e.g., by varying hybridization conditions.

"Human artificial chromosomes" (HACs), as described herein, are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size, and which contain all of the elements required for stable mitotic chromosome segregation and maintenance. (See, e.g., Harrington, J. J. et al. (1997) *Nat Genet.* 15:345-355.)

The term "humanized antibody," as used herein, refers to antibody molecules in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization," as the term is used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

As used herein, the term "hybridization complex" as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_{0t} or R_{0t} analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" or "addition," as used herein, refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, to the sequence found in the naturally occurring molecule.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be character-

ized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

The term "microarray," as used herein, refers to an array of distinct polynucleotides or oligonucleotides arrayed on a substrate, such as paper, nylon or any other type of membrane, filter, chip, glass slide, or any other suitable solid support.

The term "modulate," as it appears herein, refers to a change in the activity of EVL1. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of EVL1.

The phrases "nucleic acid" or "nucleic acid sequence," as used herein, refer to an oligonucleotide, nucleotide, polynucleotide, or any fragment thereof, to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material. In this context, "fragments" refers to those nucleic acid sequences which are greater than about 60 nucleotides in length, and most preferably are at least about 100 nucleotides, at least about 1000 nucleotides, or at least about 10,000 nucleotides in length.

The terms "operably associated" or "operably linked," as used herein, refer to functionally related nucleic acid sequences. A promoter is operably associated or operably linked with a coding sequence if the promoter controls the transcription of the encoded polypeptide. While operably associated or operably linked nucleic acid sequences can be contiguous and in reading frame, certain genetic elements, e.g., repressor genes, are not contiguously linked to the encoded polypeptide but still bind to operator sequences that control expression of the polypeptide.

The term "oligonucleotide," as used herein, refers to a nucleic acid sequence of at least about 6 nucleotides to 60 nucleotides, preferably about 15 to 30 nucleotides, and most preferably about 20 to 25 nucleotides, which can be used in PCR amplification or in a hybridization assay or microarray. As used herein, the term "oligonucleotide" is substantially equivalent to the terms "armpimers," "primers," "oligomers," and "probes," as these terms are commonly defined in the art.

"Peptide nucleic acid" (PNA), as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA and RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell. (See, e.g., Nielsen, P. E. et al. (1993) *Anticancer Drug Des.* 8:53-63.)

The term "sample," as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acids encoding EVL1, or fragments thereof, or EVL1 itself may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a solid support; a tissue; a tissue print; etc.

As used herein, the terms "specific binding" or "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, or an antagonist. The interaction is dependent upon the presence of a particular structure of the protein recognized by the binding molecule (i.e., the antigenic determinant or epitope). For example, if

an antibody is specific for epitope "A," the presence of a polypeptide containing the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

As used herein, the term "stringent conditions" refers to conditions which permit hybridization between polynucleotide sequences and the claimed polynucleotide sequences. Suitably stringent conditions can be defined by, for example, the concentrations of salt or formamide in the prehybridization and hybridization solutions, or by the hybridization temperature, and are well known in the art. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature.

For example, hybridization under high stringency conditions could occur in about 50% formamide at about 37° C. to 42° C. Hybridization could occur under reduced stringency conditions in about 35% to 25% formamide at about 30° C. to 35° C. In particular, hybridization could occur under high stringency conditions at 42° C. in 50% formamide, 5xSSPE, 0.3% SDS, and 200 µg/ml sheared and denatured salmon sperm DNA. Hybridization could occur under reduced stringency conditions as described above, but in 35% formamide at a reduced temperature of 35° C. The temperature range corresponding to a particular level of stringency can be further narrowed by calculating the purine to pyrimidine ratio of the nucleic acid of interest and adjusting the temperature accordingly. Variations on the above ranges and conditions are well known in the art.

The term "substantially purified," as used herein, refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably about 75% free, and most preferably about 90% free from other components with which they are naturally associated.

A "substitution," as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation," as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed" cells includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, and refers to cells which transiently express the inserted DNA or RNA for limited periods of time.

A "variant" of EVL1, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (e.g., replacement of leucine with isoleucine). More rarely, a variant may have "nonconservative" changes (e.g., replacement of glycine with tryptophan). Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

The Invention

The invention is based on the discovery of a new human ena/VASP-like protein splice variant (EVL1), the polynucleotides encoding EVL1, and the use of these compositions for the diagnosis, treatment, or prevention of reproductive, immunological, vesicle trafficking, nervous system, developmental, and neoplastic disorders.

Nucleic acids encoding the EVL1 of the present invention were first identified in Incyte Clone 3089412 from the aorta cDNA library (HEAONOT03) using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 3089412 (HEAONOT03), 2836864 (TYMNOT03), 1822064 (GBLATUT01), 1446806 (PLACNOT02), 1556238 (BLADTUT04), 1209813 (BRSTNOT02), and the shotgun sequence SAEA02787.

In one embodiment, the invention encompasses a polypeptide consisting of the amino acid sequence of SEQ ID NO:1, as shown in FIGS. 1A, 1B, 1C, 1D, and 1E. EVL1 is 418 amino acids in length and has two potential N-glycosylation sites at residues N64, and N319; one potential cAMP- and cGMP-dependent protein kinase phosphorylation site at residue S160; eight potential casein kinase II phosphorylation sites at residues S96, S132, S215, S216, S253, S285, S298, and S371; six potential protein kinase C phosphorylation sites at residues T21, S22, T48, S123, T252, and S287; and a proline cluster from about residue P346 to about residue P368 and a predicted turn-coil-turn three-fold repeat structure from about residue T345 to about residue D374, indicative of a protein-protein interacting domain. As shown in FIGS. 2A, 2B, and 2C, EVL1 has chemical and structural homology with mouse ena/VASP-like protein (G1644453; SEQ ID NO:3), and human VASP (GI 624964; SEQ ID NO:4). In particular, EVL1 and mouse ena/VASP-like protein share 92% identity, two potential N-glycosylation sites, one potential cAMP- and cGMP-dependent protein kinase phosphorylation site, seven potential casein kinase II phosphorylation sites, and six potential protein kinase C phosphorylation sites. In addition, EVL1 and mouse ena/VASP-like protein have similar isoelectric points, 9.2 and 8.7, respectively. As illustrated by FIGS. 3A and 3B, EVL1 and mouse ena/VASP-like protein have rather similar hydrophobicity plots. The fragment of SEQ ID NO:2 from about nucleotide 1161 to about nucleotide 1197 is useful for designing oligonucleotides or to be used directly as a hybridization probe. Northern analysis shows the expression of this sequence in various libraries, at least 49% of which are immortalized or cancerous and at least 26% of which involve immune response. Of particular note is the expression of EVL1 in gastrointestinal, cardiovascular, neural, and developmental tissue; and in prostate, breast, ovary, and uterus tissue.

The invention also encompasses EVL1 variants. A preferred EVL1 variant is one which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% amino acid sequence identity to the EVL1 amino acid sequence, and which contains at least one functional or structural characteristic of EVL1.

The invention also encompasses polynucleotides which encode EVL1. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising the sequence of SEQ ID NO:2, which encodes an EVL1.

The invention also encompasses a variant of a polynucleotide sequence encoding EVL1. In particular, such a variant polynucleotide sequence will have at least about 80%, more

preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding EVL1. A particular aspect of the invention encompasses a variant of SEQ ID NO:2 which has at least about 80%, more preferably at least about 90%, and most preferably at least about 95% polynucleotide sequence identity to SEQ ID NO:2. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of EVL1.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding EVL1, some bearing minimal homology to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring EVL1, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode EVL1 and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring EVL1 under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding EVL1 or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding EVL1 and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode EVL1 and EVL1 derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding EVL1 or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:2, or a fragment of SEQ ID NO:2, under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399-407; and Kimmel, A. R. (1987) *Methods Enzymol.* 152:507-511.)

Methods for DNA sequencing are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical Corp., Cleveland, Ohio.), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, Ill.), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (GIBCO/BRL, Gaithersburg, Md.). Preferably, the process is automated with machines such as the microlab 2200 (Hamilton, Reno, Nev.), Peltier Thermal-Cycler (PTC200; MJ Research, Watertown, Mass.) and the ABI catalyst and 373 and 377 DNA sequencers (Perkin Elmer).

The nucleic acid sequences encoding EVL1 may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.) In particular, genomic DNA is first amplified in the presence of a primer complementary to a linker sequence within the vector and a primer specific to the region predicted to encode the gene. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) The primers may be designed using commercially available software such as OLIGO 4.06 primer analysis software (National Biosciences Inc., Plymouth, Minn.) or another appropriate program to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68° C. to 72° C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR, which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.) In this method, multiple restriction enzyme digestions and ligations may be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J. D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060.) Additionally, one may use PCR, nested primers, and PROMOTER FINDER libraries to walk genomic DNA (Clontech, Palo Alto, Calif.). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable in that they will include more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Perkin Elmer), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially pref-

erable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode EVL1 may be used in recombinant DNA molecules to direct expression of EVL1, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express EVL1.

As will be understood by those of skill in the art, it may be advantageous to produce EVL1-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter EVL1-encoding sequences for a variety of reasons including, but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding EVL1 may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of EVL1 activity, it may be useful to encode a chimeric EVL1 protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the EVL1 encoding sequence and the heterologous protein sequence, so that EVL1 may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding EVL1 may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M. H. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 215-223, and Hom, T. et al. (1980) *Nucl. Acids Res. Symp. Ser.* 225-232.) Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of EVL1, or a fragment thereof. For example, peptide synthesis can be performed using various solid-phase techniques. (See, e.g., Roberge, J. Y. et al. (1995) *Science* 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Perkin Elmer).

The newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R. M. and F. Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, T. (1983) *Proteins, Structures and Molecular Properties*, WH Freeman and Co., New York, N.Y.) Additionally, the amino acid sequence of EVL1, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active EVL1, the nucleotide sequences encoding EVL1 or derivatives thereof may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding EVL1 and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) *Molecular Cloning. A Laboratory Manual*, Cold Spring Harbor Press, Plainview, N.Y., ch. 4, 8, and 16-17; and Ausubel, F. M. et al. (1995, and periodic supplements) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, N.Y., ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding EVL1. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus (CaMV) or tobacco mosaic virus (TMV)) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions, e.g., enhancers, promoters, and 5' and 3' untranslated regions, of the vector and polynucleotide sequences encoding EVL1 which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters, e.g., hybrid lacZ promoter of the BLUESCRIPT phagemid (Stratagene, La Jolla, Calif.) or PSPOT1 plasmid (GIBCO/BRL), may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO, and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding EVL1, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for EVL1. For example, when large quantities of EVL1 are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, multi-functional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the sequence encoding EVL1 may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of β -galactosidase so that a hybrid protein is produced, and pIN vectors. (See, e.g., Van Heeke, G. and S. M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509.) pGEX vectors (Pharmacia Biotech, Uppsala, Sweden) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion

proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH, may be used. (See, e.g., Ausubel, supra; and Grant et al. (1987) *Methods Enzymol.* 153:516-544.) In cases where plant expression vectors are used, the expression of sequences encoding EVL1 may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV. (Takamatsu, N. (1987) *EMBO J.* 6:307-311.) Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews. (See, e.g., Hobbs, S. or Murry, L. E. in *McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York, N.Y.; pp. 191-196.) An insect system may also be used to express EVL1. For example, in one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequences encoding EVL1 may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of sequences encoding EVL1 will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which EVL1 may be expressed. (See, e.g., Engelhard, E. K. et al. (1994) *Proc. Nat. Acad. Sci.* 91:3224-3227.)

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding EVL1 may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing EVL1 in infected host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding EVL1. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding EVL1 and its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in

cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probi. Cell Differ.* 20:125-162.)

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding, and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and W138), are available from the American Type Culture Collection (ATCC, Bethesda, Md.) and may be chosen to ensure the correct modification and processing of the foreign protein.

For long term, high yield production of recombinant proteins, stable expression is preferred. For example, cell lines capable of stably expressing EVL1 can be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase genes and adenine phosphoribosyltransferase genes, which can be employed in tk or apr cells, respectively. (See, e.g., Wigler, M. et al. (1977) *Cell* 11:223-232; and Lowy, I. et al. (1980) *Cell* 22:817-823) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, dhfr, confers resistance to methotrexate; npt confers resistance to the aminoglycosides neomycin and G-418; and als or pat confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci.* 77:3567-3570; Colbere-Garapin, F. et al (1981) *J. Mol. Biol.* 150:1-14; and Murry, supra.) Additional selectable genes have been described, e.g., trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine. (See, e.g., Hartman, S. C. and R. C. Mulligan (1988) *Proc. Natl. Acad. Sci.* 85:8047-8051.) Recently, the use of visible markers has gained popularity with such markers as anthocyanins, B glucuronidase and its substrate GUS, luciferase and its substrate luciferin. Green fluorescent proteins (GFP) (Clontech, Palo Alto, Calif.) are also used (See, e.g., Chalfie, M. et al. (1994) *Science* 263:802-805.) These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C. A. et al. (1995) *Methods Mol. Biol.* 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding EVL1 is inserted within a marker gene sequence, transformed cells containing sequences encoding EVL1 can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding EVL1 under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding EVL1 and express EVL1 may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

The presence of polynucleotide sequences encoding EVL1 can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding EVL1. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding EVL1 to detect transformants containing DNA or RNA encoding EVL1.

A variety of protocols for detecting and measuring the expression of EVL1, using either polyclonal or monoclonal antibodies specific for the protein, are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on EVL1 is preferred, but a competitive binding assay may be employed. These and other assays are well described in the art. (See, e.g., Hampton, R. et al. (1990) *Serological Methods, a Laboratory Manual*, APS Press, St Paul, Minn., Section IV; and Maddox, D. E. et al. (1983) *J. Exp. Med.* 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding EVL1 include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding EVL1, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Pharmacia & Upjohn (Kalamazoo, Minn.), Promega (Madison, Wis.), and U.S. Biochemical Corp. (Cleveland, Ohio.). Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding EVL1 may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The

protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode EVL1 may be designed to contain signal sequences which direct secretion of EVL1 through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding EVL1 to nucleotide sequences encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, Wash.). The inclusion of cleavable linker sequences, such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, Calif.), between the purification domain and the EVL1 encoding sequence may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing EVL1 and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on immobilized metal ion affinity chromatography. (IMAC) (See, e.g., Porath, J. et al. (1992) *Prot. Exp. Purif.* 3: 263-281.) The enterokinase cleavage site provides a means for purifying EVL1 from the fusion protein. (See, e.g., Kroll, D. J. et al. (1993) *DNA Cell Biol.* 12:441-453.) Fragments of EVL1 may be produced not only by recombinant production, but also by direct peptide synthesis using solid-phase techniques. (See, e.g., Creighton, T.E. (1984) *Protein: Structures and Molecular Properties*, pp. 55-60, W. H. Freeman and Co., New York, N.Y.) Protein synthesis may be performed by manual techniques or by automation. Automated synthesis may be achieved, for example, using the Applied Biosystems 431 A peptide synthesizer (Perkin Elmer). Various fragments of EVL1 may be synthesized separately and then combined to produce the full length molecule.

Therapeutics

Chemical and structural homology exists among EVL1, mouse ena/VASP-like protein (GI 1644453), and human VASP (GI 624964). In addition, EVL1 is expressed in cancer and the immune response; in gastrointestinal, cardiovascular, neural, and developmental tissue; and in prostate, breast, ovary, and uterus tissue. Therefore, EVL1 appears to play a role in reproductive, immunological, vesicle trafficking, nervous system, developmental, and neoplastic disorders.

Therefore, in one embodiment, EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a reproductive disorder. Such reproductive disorders can include, but are not limited to, disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, autoimmune disorders, ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, and prostatitis, carcinoma of the male breast and gynecocarcinoma.

In another embodiment, a vector capable of expressing EVL1 or a fragment or derivative thereof may be adminis-

tered to a subject to treat or prevent a reproductive disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified EVL1 in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a reproductive disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of EVL1 may be administered to a subject to treat or prevent a reproductive disorder including, but not limited to, those listed above.

In another embodiment, EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent an immunological disorder. Such immunological disorders can include, but are not limited to, AIDS, Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, ulcerative colitis, Werner syndrome, and complications of cancer, hemodialysis, and extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections; and trauma.

In another embodiment, a vector capable of expressing EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent an immunological disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified EVL1 in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent an immunological disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of EVL1 may be administered to a subject to treat or prevent an immunological disorder including, but not limited to, those listed above.

In another embodiment, EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a vesicle trafficking disorder. Such vesicle trafficking disorders can include, but are not limited to, cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, diabetes mellitus, diabetes insipidus, hyper- and hypoglycemia, Grave's disease, goiter, Cushing's disease, and Addison's disease; gastrointestinal disorders including ulcerative colitis, gastric and duodenal ulcers; other conditions associated with abnormal vesicle trafficking including AIDS; allergies including hay fever, asthma, and urticaria (hives); autoimmune hemolytic anemia; proliferative glomerulonephritis; inflammatory bowel disease; multiple sclerosis; myasthenia gravis; rheumatoid and osteoarthritis; scleroderma; Chediak-Higashi and Sjögren's syndromes; systemic lupus erythematosus; toxic shock syndrome; traumatic tissue damage; and viral, bacterial, fungal, helminth, and protozoal infections.

In another embodiment, a vector capable of expressing EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a vesicle trafficking disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified EVL1 in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a vesicle trafficking disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of EVL1 may be administered to a subject to treat or prevent a vesicle trafficking disorder including, but not limited to, those listed above.

In another embodiment, EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a nervous system disorder. Such nervous system disorders can include, but are not limited to, akathisia, Alzheimer's disease, amnesia, amyotrophic lateral sclerosis, bipolar disorder, catatonia, cerebral neoplasms, dementia, depression, diabetic neuropathy, Down's syndrome, tardive dyskinesia, dystonias, epilepsy, Huntington's disease, multiple sclerosis, neurofibromatosis, Parkinson's disease, paranoid psychoses, postherpetic neuralgia, schizophrenia, and Tourette's disorder; angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, and pheochromocytoma.

In another embodiment, a vector capable of expressing EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a nervous system disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified EVL1 in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a nervous system disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of EVL1 may be administered to a subject to treat or prevent a nervous system disorder including, but not limited to, those listed above.

In another embodiment, EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a developmental disorder. The term "developmental disorder" refers to any disorder associated with development or function of a tissue, organ, or system of a subject (such as the brain, adrenal gland, kidney, skeletal or reproductive system). Such developmental disorders can include, but are not limited to, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome, Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spinal bifida, and congenital glaucoma, cataract, or sensorineural hearing loss.

In another embodiment, a vector capable of expressing EVL1 or a fragment or derivative thereof may be administered to a subject to treat or prevent a developmental disorder including, but not limited to, those described above.

In a further embodiment, a pharmaceutical composition comprising a substantially purified EVL1 in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a developmental disorder including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of EVL1 may be administered to a subject to treat or prevent a developmental disorder including, but not limited to, those listed above.

In a further embodiment, an antagonist of EVL1 may be administered to a subject to treat or prevent a neoplastic disorder. Such a neoplastic disorder may include, but is not limited to, adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, an antibody which specifically binds EVL1 may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express EVL1.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding EVL1 may be administered to a subject to treat or prevent a neoplastic disorder including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of EVL1 may be produced using methods which are generally known in the art. In particular, purified EVL1 may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind EVL1. Antibodies to EVL1 may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with EVL1 or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysocleithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guérin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to EVL1 have an amino acid sequence consisting of at least about 5 amino acids, and, more preferably, of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein and contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of EVL1 amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to EVL1 may be prepared using any technique which provides for the production of antibody

molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R. J. et al. (1983) *Proc. Natl. Acad. Sci.* 80:2026-2030; and Cole, S. P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S. L. et al. (1984) *Proc. Natl. Acad. Sci.* 81:6851-6855; Neuberger, M. S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce EVL1-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotype composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton D. R. (1991) *Proc. Natl. Acad. Sci.* 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci.* 86: 3833-3837; and Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for EVL1 may also be generated. For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W. D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between EVL1 and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering EVL1 epitopes is preferred, but a competitive binding assay may also be employed. (Maddox, supra.)

In another embodiment of the invention, the polynucleotides encoding EVL1, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding EVL1 may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding EVL1. Thus, complementary molecules or fragments may be used to modulate EVL1 activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding EVL1.

Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide

sequences to the targeted organ, tissue, or cell population. Methods which are well known to those skilled in the art can be used to construct vectors which will express nucleic acid sequences complementary to the polynucleotides of the gene encoding EVL1. (See, e.g., Sambrook, supra; and Ausubel, supra.)

Genes encoding EVL1 can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide, or fragment thereof, encoding EVL1. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector, and may last even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5', or regulatory regions of the gene encoding EVL1. Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J. E. et al. (1994) in Huber, B. E. and B. I. Carr, *Molecular and Immunologic Approaches*, Futura Publishing Co., Mt. Kisco, N.Y., pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding EVL1.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding EVL1. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C. K. et al. (1997) *Nature Biotechnology* 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical or sterile composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of EVL1, antibodies to EVL1, and mimetics, agonists, antagonists, or inhibitors of EVL1. The compositions may be administered alone or in combination with at least one other agent, such as a stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs, or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of *Remington's Pharmaceutical Sciences* (Maack Publishing Co., Easton, Pa.).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combining active compounds with solid excipient and processing the resultant mixture of granules (optionally, after grinding) to obtain tablets or dragee cores. Suitable auxiliaries can be added, if desired. Suitable excipients include carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, and sorbitol; starch

from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums, including arabic and tragacanth; and proteins, such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, and alginic acid or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils, such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate, triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents to increase the solubility of the compounds and allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acids. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1 mM to 50 mM histidine, 0.1% to 2% sucrose, and 2% to 7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of EVL1, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredi-

ents are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells or in animal models such as mice, rats, rabbits, dogs, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example EVL1 or fragments thereof, antibodies of EVL1, and agonists, antagonists or inhibitors of EVL1, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED50 (the dose therapeutically effective in 50% of the population) or LD50 (the dose lethal to 50% of the population) statistics. The dose ratio of therapeutic to toxic effects is the therapeutic index, which can be expressed as the ED50/LD50 ratio. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μ g to 100,000 μ g, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Diagnostics

In another embodiment, antibodies which specifically bind EVL1 may be used for the diagnosis of disorders characterized by expression of EVL1, or in assays to monitor patients being treated with EVL1 or agonists, antagonists, or inhibitors of EVL1. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for EVL1 include methods which utilize the antibody and a label to detect EVL1 in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring EVL1, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of EVL1 expression. Normal or standard values for EVL1 expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with is antibody to EVL1 under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, preferably by photometric means. Quantities of EVL1 expressed in subject, samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding EVL1 may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of EVL1 may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of EVL1, and to monitor regulation of EVL1 levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding EVL1 or closely related molecules may be used to identify nucleic acid sequences which encode EVL1. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low), will determine whether the probe identifies only naturally occurring sequences encoding EVL1, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the EVL1 encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:2 or from genomic sequences including promoters, enhancers, and introns of the EVL1 gene.

Means for producing specific hybridization probes for DNAs encoding EVL1 include the cloning of polynucleotide sequences encoding EVL1 or EVL1 derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as 32 P or 35 S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding EVL1 may be used for the diagnosis of a disorder associated with expression of EVL1. Examples of such a disorder include, but are not limited to, a reproductive disorder, such as, disorders of prolactin production; infertility, including tubal disease, ovulatory defects, and endometriosis; disruptions of the estrous cycle, disruptions of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, autoimmune disorders, ectopic pregnancy, and teratogenesis; cancer of the breast, fibrocystic breast disease, and galactorrhea; disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia,

and prostatitis, carcinoma of the male breast and gynecoma-
stia; an immunological disorder, such as, AIDS, Addison's
disease, adult respiratory distress syndrome, allergies,
ankylosing spondylitis, amyloidosis, anemia, asthma,
atherosclerosis, autoimmune hemolytic anemia, autoimmune
thyroiditis, bronchitis, cholecystitis, contact dermatitis,
Crohn's disease, atopic dermatitis, dermatomyositis, diabetes
mellitus, emphysema, erythema nodosum, atrophic gastritis,
glomerulonephritis, Goodpasture's syndrome, gout, Graves'
disease, Hashimoto's thyroiditis, hypereosinophilia, irritable
bowel syndrome, lupus erythematosus, multiple sclerosis,
myasthenia gravis, myocardial or pericardial inflammation,
osteoarthritis, osteoporosis, pancreatitis, polymyositis,
rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic
anaphylaxis, systemic lupus erythematosus, systemic sclerosis,
ulcerative colitis, Werner syndrome, and complications of
cancer, hemodialysis, and extracorporeal circulation; viral,
bacterial, fungal, parasitic, protozoal, and helminthic
infections; and trauma; a vesicle trafficking disorder, such
as, cystic fibrosis, glucose-galactose malabsorption syndrome,
hypercholesterolemia, diabetes mellitus, diabetes insipidus,
hyper- and hypoglycemia, Grave's disease, goiter, Cushing's
disease, and Addison's disease; gastrointestinal disorders
including ulcerative colitis, gastric and duodenal ulcers;
other conditions associated with abnormal vesicle trafficking
including AIDS; allergies including hay fever, asthma, and
urticaria (hives); autoimmune hemolytic anemia; proliferative
glomerulonephritis; inflammatory bowel disease; multiple
sclerosis; myasthenia gravis; rheumatoid and osteoarthritis;
scleroderma; Chediak-Higashi and Sjögren's syndromes;
systemic lupus erythematosus; toxic shock syndrome;
traumatic tissue damage; and viral, bacterial, fungal,
helminth, and protozoal infections; a nervous system
disorder, such as, akathisia, Alzheimer's disease, amnesia,
amyotrophic lateral sclerosis, bipolar disorder, catatonia,
cerebral neoplasms, dementia, depression, diabetic
neuropathy, Down's syndrome, tardive dyskinesia,
dystonias, epilepsy, Huntington's disease, multiple
sclerosis, neurofibromatosis, Parkinson's disease, paranoid
psychoses, postherpetic neuralgia, schizophrenia, and
Tourette's disorder; angina, anaphylactic shock, arrhythmias,
asthma, cardiovascular shock, Cushing's syndrome,
hypertension, hypoglycemia, myocardial infarction,
migraine, and pheochromocytoma; a developmental
disorder, such as, renal tubular acidosis, anemia, Cushing's
syndrome, achondroplastic dwarfism, Duchenne and Becker
muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR
syndrome, Smith-Magenis syndrome, myelodysplastic
syndrome, hereditary mucocutaneous dysplasia, hereditary
keratodermas, hereditary neuropathies such as Charcot-Marie-
Tooth disease and neurofibromatosis, hypothyroidism,
hydrocephalus, seizure disorders such as Sydenham's chorea
and cerebral palsy, spinal bifida, and congenital glaucoma,
cataract, or sensorineural hearing loss, and a neoplastic
disorder, such as, adenocarcinoma, leukemia, lymphoma,
melanoma, myeloma, sarcoma, teratocarcinoma, and, in
particular, cancers of the adrenal gland, bladder, bone,
bone marrow, brain, breast, cervix, gall bladder, ganglia,
gastrointestinal tract, heart, kidney, liver, lung, muscle,
ovary, pancreas, parathyroid, penis, prostate, salivary
glands, skin, spleen, testis, thymus, thyroid, and uterus.
The polynucleotide sequences encoding EVL1 may be used in
Southern or northern analysis, dot blot, or other membrane-
based technologies; in PCR technologies; in dipstick, pin,
and ELISA assays; and in microarrays utilizing fluids or
tissues from patients to detect altered EVL1 expres-

sion. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding EVL1 may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding EVL1 may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding EVL1 in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of EVL1, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding EVL1, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding EVL1 may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding EVL1, or a fragment of a polynucleotide complementary to the polynucleotide encoding EVL1, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of EVL1 include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P. C. et al. (1993) *J. Immunol. Methods* 159:235-244; and Duplaa, C. et al. (1993) *Anal. Biochem* 212:229-236.) The

speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as targets in a microarray. The microarray can be used to monitor the expression level of large numbers of genes simultaneously and to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, and to develop and monitor the activities of therapeutic agents.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T. M. et al. (1995) U.S. Pat. No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R. A. et al. (1997) Proc. Natl. Acad. Sci. 94:2150-2155; and Heller, M. J. et al. (1997) U.S. Pat. No. 5,605,662.)

In another embodiment of the invention, nucleic acid sequences encoding EVL1 may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial plasmid constructions, or single chromosome cDNA libraries. (See, e.g., Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B. J. (1991) Trends Genet. 7:149-154.)

Fluorescent in situ hybridization (FISH) may be correlated with other physical chromosome mapping techniques and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, R. A. (ed.) *Molecular Biology and Biotechnology*, VCH Publishers New York, N.Y., pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) site. Correlation between the location of the gene encoding EVL1 on a physical chromosomal map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder. The nucleotide sequences of the invention may be used to detect differences in gene sequences among normal, carrier, and affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, e.g., AT to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R. A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the subject invention may also be used to detect

differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, EVL1, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between EVL1 and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application W084/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with EVL1, or fragments thereof, and washed. Bound EVL1 is then detected by methods well known in the art. Purified EVL1 can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding EVL1 specifically compete with a test compound for binding EVL1. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with EVL1.

In additional embodiments, the nucleotide sequences which encode EVL1 may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

EXAMPLES

1. HEAONOT03 cDNA Library Construction

The HEAONOT03 cDNA library was constructed from normal aorta tissue obtained from a 27-year-old Caucasian female who died from intracranial bleeding.

The frozen tissue was homogenized and lysed TRIZOL reagent (1 g tissue/10 ml TRIZOL, Catalog #10296-028; GIBCO-BRL), a monophasic solution of phenol and guanidine isothiocyanate, using a Polytron PT-3000 homogenizer (Brinkmann Instruments, Westbury, N.Y.). After a brief incubation on ice, chloroform was added (1:5 v/v) and the lysate was centrifuged. The upper aqueous layer was removed to a fresh tube and the RNA precipitated with isopropanol, resuspended in DEPC-treated water, and treated with DNase for 25 min at 37° C. The RNA was extracted and precipitated as described before. The mRNA was then isolated using the OLIGOTEX (QIAGEN, Inc., Chatsworth, Calif.) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SUPERScript plasmid system (Catalog #18248-013, GIBCO-BRL). cDNA synthesis was initiated with a Not I-oligo d(T) primer. Double-stranded cDNA was blunted, ligated to EcoR I adaptors, digested with Not I, fractionated on a SEPHAROSE CL4B column (Catalog

#275105-01; Pharmacia), and those cDNAs exceeding 400 bp were ligated into the Not I- and EcoR I sites of the pINCY 1 vector (Incyte). The plasmid pINCY 1 was subsequently transformed into DH5 α competent cells (Catalog #18258-012; GIBCO-BRL).

II. Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using the REAL PREP 96 plasmid kit (Catalog #26173; QIAGEN, Inc.). The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711, GIBCO-BRL) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4° C.

The cDNAs were sequenced by the method of Sanger et al. (1975, *J. Mol. Biol.* 94:441f), using a MICROLAB 2200 (Hamilton, Reno, Nev.) in combination with Peltier thermal cyclers (PTC200 from MJ Research, Watertown, Mass.) and Applied Biosystems 377 DNA sequencing systems, and the reading frame was determined.

III. Homology Searching of cDNA Clones and Their Deduced Proteins

The nucleotide sequences and/or amino acid sequences of the Sequence Listing were used to query sequences in the GenBank, SwissProt, BLOCKS, and Pima II databases. These databases, which contain previously identified and annotated sequences, were searched for regions of homology using BLAST (Basic Local Alignment Search Tool). (See, e.g., Altschul, S. F. (1993) *J. Mol. Evol.* 36:290-300; and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410.)

BLAST produced alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST was especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Other algorithms could have been used when dealing with primary sequence patterns and secondary structure gap penalties. (See, e.g., Smith, T. et al. (1992) *Protein Engineering* 5:35-51.) The sequences disclosed in this application have lengths of at least 49 nucleotides and have no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach searched for matches between a query sequence and a database sequence. BLAST evaluated the statistical significance of any matches found, and reported only those matches that satisfy the user-selected threshold of significance. In this application, threshold was set at 10⁻²⁵ for nucleotides and 10⁻⁸ for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for primate PF-0471 US (pri), rodent (rod), and other mammalian sequences (mam), produced amino acid sequences from the same clones were then searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp), for homology.

Additionally, sequences identified from cDNA libraries may be analyzed to identify those gene sequences encoding

conserved protein motifs using an appropriate analysis program, e.g., the Block 2 Bioanalysis Program (Incyte, Palo Alto, Calif.). This motif analysis program, based on sequence information contained in the Swiss-Prot-Database and PROSITE, is a method of determining the function of uncharacterized proteins translated from genomic or cDNA sequences. (See, e.g., Bairoch, A. et al. (1997) *Nucleic Acids Res.* 25:217-221; and Attwood, T. K. et al. (1997) *J. Chem. Inf. Comput. Sci.* 37:417424.) PROSITE may be used to identify common functional or structural domains in divergent proteins. The method is based on weight matrices. Motifs identified by this method are then calibrated against the SWISS-PROT database in order to obtain a measure of the chance distribution of the matches.

In another alternative, Hidden Markov models (HMMs) may be used to find protein domains, each defined by a dataset of proteins known to have a common biological function. (See, e.g., Pearson, W. R. and D. J. Lipman (1988) *Proc. Natl. Acad. Sci.* 85:2444-2448; and Smith, T. F. and M. S. Waterman (1981) *J. Mol. Biol.* 147:195-197.) HMMs were initially developed to examine speech recognition patterns, but are now being used in a biological context to analyze protein and nucleic acid sequences as well as to model protein structure. (See, e.g., Krogh, A. et al. (1994) *J. Mol. Biol.* 235:1501-1531; and Collin, M. et al. (1993) *Protein Sci.* 2:305-314.) HMMs have a formal probabilistic basis and use position-specific scores for amino acids or nucleotides. The algorithm continues to incorporate information from newly identified sequences to increase its motif analysis capabilities.

IV. Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; and Ausubel, F. M. et al. supra, ch. 4 and 16.)

Analogous computer techniques applying BLAST are used to search for identical or related molecules in nucleotide databases such as GENBANK or LIFESEQ database (Incyte Pharmaceuticals). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score, which is defined as:

$$\frac{\% \text{ sequence identity} \times \% \text{ maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1% to 2% error, and, with a product score of 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analysis are reported as a list of libraries in which the transcript encoding EVL1 occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

V. Extension of EVL1 Encoding Polynucleotides

The nucleic acid sequence of Incyte Clone 3089412 was used to design oligonucleotide primers for extending a partial nucleotide sequence to full length. One primer was synthesized to initiate extension of an antisense polynucleotide, and the other was synthesized to initiate extension of a sense polynucleotide. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 software (National Biosciences, Plymouth, Minn.), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68° C. to about 72° C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (GIBCO/BRL) were used to extend the sequence. If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. PCR was performed using the Peltier thermal cycler (PTC200; M. J. Research, Watertown, Mass.), beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, with the following parameters:

Step 1	94° C. for 1 min (initial denaturation)
Step 2	65° C. for 1 min
Step 3	68° C. for 6 min
Step 4	94° C. for 15 sec
Step 5	65° C. for 1 min
Step 6	68° C. for 7 min
Step 7	Repeat steps 4 through 6 for an additional 15 cycles
Step 8	94° C. for 15 sec
Step 9	65° C. for 1 min
Step 10	68° C. for 7:15 min
Step 11	Repeat steps 8 through 10 for an additional 12 cycles
Step 12	72° C. for 8 min
Step 13	4° C. (and holding)

A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6% to 0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQUICK (QIAGEN Inc., Chatsworth, Calif.), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13 μ l of ligation buffer, 1 μ l T4-DNA ligase (15 units) and 1 μ l T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2 to 3 hours, or overnight at 16° C. Competent *E. coli* cells (in 40 μ l of appropriate media) were transformed with 3 μ l of ligation mixture and cultured in 80 μ l of SOC medium. (See, e.g., Sambrook, supra, Appendix A, p. 2.) After incubation for one hour at 37° C., the *E. coli* mixture was plated on Luria Bertani (LB) agar (See, e.g., Sambrook, supra, Appendix A, p. 1) containing 2xCarb. The following day, several colonies were randomly picked from each plate and cultured in 150 μ l of liquid LB/2xCarb medium placed in an individual well of an appropriate commercially-available sterile 96-well

microtiter plate. The following day, 5 μ l of each overnight culture was transferred into a non-sterile 96-well plate and, after dilution 1:10 with water, 5 μ l from each sample was transferred into a PCR array.

For PCR amplification, 18 μ l of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

Step 1	94° C. for 60 sec
Step 2	94° C. for 20 sec
Step 3	55° C. for 30 sec
Step 4	72° C. for 90 sec
Step 5	Repeat steps 2 through 4 for an additional 29 cycles
Step 6	72° C. for 180 sec
Step 7	4° C. (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequence of SEQ ID NO:2 is used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

VI. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:2 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ -³²P] adenosine triphosphate (Amersham, Chicago, Ill.), and T4 polynucleotide kinase (DuPont NEN, Boston, Mass.). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine resin column (Pharmacia & Upjohn, Kalamazoo, Mich.). An aliquot containing 10⁷ counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN, Boston, Mass.).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, N.H.). Hybridization is carried out for 16 hours at 40° C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMATAR film (Kodak, Rochester, N.Y.) is exposed to the blots, hybridization patterns are compared visually.

VII. Microarrays

A chemical coupling procedure and an ink jet device can be used to synthesize array elements on the surface of a substrate. (See, e.g., Baldeschweiler, supra.) An array analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV,

chemical, or mechanical bonding procedures. A typical array may be produced by hand or using available methods and machines and contain any appropriate number of elements. After hybridization, nonhybridized probes are removed and a scanner used to determine the levels and patterns of fluorescence. The degree of complementarity and the relative abundance of each probe which hybridizes to an element on the microarray may be assessed through analysis of the scanned images.

Full-length cDNAs, Expressed Sequence Tags (ESTs), or fragments thereof may comprise the elements of the microarray. Fragments suitable for hybridization can be selected using software well known in the art such as LASERGENE. Full-length cDNAs, ESTs, or fragments thereof corresponding to one of the nucleotide sequences of the present invention, or selected at random from a cDNA library relevant to the present invention, are arranged on an appropriate substrate, e.g., a glass slide. The cDNA is fixed to the slide using, e.g., UV cross-linking followed by thermal and chemical treatments and subsequent drying. (See, e.g., Schena, M. et al. (1995) *Science* 270:467-470; and Shalon, D. et al. (1996) *Genome Res.* 6:639-645.) Fluorescent probes are prepared and used for hybridization to the elements on the substrate. The substrate is analyzed by procedures described above.

VIII. Complementary Polynucleotides

Sequences complementary to the EVL1-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring EVL1. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software and the coding sequence of EVL1. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the EVL1 -encoding transcript.

IX. Expression of EVL1

Expression of EVL1 is accomplished by subcloning the cDNA into an appropriate vector and transforming the vector into host cells. This vector contains an appropriate promoter, e.g., β -galactosidase upstream of the cloning site, operably associated with the cDNA of interest. (See, e.g., Sambrook, supra pp. 404433; and Rosenberg, M. et al. (1983) *Methods Enzymol.* 101:123-138.)

Induction of an isolated, transformed bacterial strain with isopropyl beta-D-thiogalactopyranoside (IPTG) using standard methods produces a fusion protein which consists of the first 8 residues of β -galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of EVL1 into bacterial growth media which can be used directly in the following assay for activity.

X. Demonstration of EVL1 Activity

The assay for human ena/VASP-like protein splice variant (EVL1) is based upon the binding affinity of mouse ena/VASP for the neural protein FE65. (Ermeikova, K. S. et al. supra.) Monkey cell line COS-7 (ATCC CRL 1651) cells are grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin mixture in 5% CO₂ atmosphere at 37° C.

3x10⁶ cells are transfected by electroporation with 20 μ g of CMV-hemagglutinin-FE65 plasmid and 20 μ g of CMV-EVL1 plasmid as known in the art. 62 h after the transfection, the cells are harvested in ice-cold PBS and centrifuged at 2000 rpm at 4° C., and the pellet is dissolved in lysis buffer (10 mM Tris HCl, pH 7.5; 150 mM NaCl; 0.1 mM sodium vanadate; 50 mM NaF; 0.5% Nonidet P-40; 1 mM phenylmethylsulfonyl fluoride; 10 μ g/ml each of aprotinin, leupeptin, and pepstatin). The extracts are clarified by centrifugation at 16,000 g at 4° C., and 4 mg of supernatant are incubated for 1 h at 4° C. with an anti-hemagglutinin monoclonal antibody or with an unrelated monoclonal antibody. Thereafter, 30 μ l of protein A-SEPHAROSE resin (Pharmacia) are added to each sample of the extract-antibody mixture, and the immunocomplexes are eluted with 50 mM Tris HCl, pH 6.8, 2% SDS, 10% glycerol, 100 mM dithiothreitol, 0.01% bromophenol blue. The proteins are resolved by 7.5% SDS-PAGE and transferred to Immobilon-P membranes (Millipore). The filter is blocked in 5% nonfat dry milk in tris-buffered saline, 0.5% Tween, buffer (TBS-T) and incubated with anti-EVL1 antibodies at 1:1000 dilution for 1 h at room temperature. After washing in TBS-T, the filter is exposed to horseradish peroxidase-conjugated protein A (Amersham Corp.) at a dilution of 1:5000 for 30 min at room temperature. The signals are detected by chemiluminescence using the ECL system (Amersham Corp.). The signal response is proportional to the activity of EVL1 in the preparation.

XI. Production of EVL1 Specific Antibodies

EVL1 substantially purified using PAGE electrophoresis (see, e.g., Harrington, M. G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. The EVL1 amino acid sequence is analyzed using DNASTAR software (DNASTAR Inc) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art. (See, e.g., Ausubel et al., ch. 11.)

Typically, the oligopeptides are 15 residues in length, and are synthesized using an Applied Biosystems 431A Peptide Synthesizer using fmoc-chemistry and coupled to KLH (Sigma, St. Louis, Mo.) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel et al. supra.) Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XII. Purification of Naturally Occurring EVL1 Using Specific Antibodies

Naturally occurring or recombinant EVL1 is substantially purified by immunoaffinity chromatography using antibodies specific for EVL1. An immunoaffinity column is constructed by covalently coupling anti-EVL1 antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing EVL1 are passed over the immunoaffinity column, and the column is washed under conditions

that allow the preferential absorbance of EVL1 (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/EVL1 binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and EVL1 is collected.

XIII. Identification of Molecules Which Interact with EVL1

EVL1, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent. (See, e.g., Bolton et al. (1973) *Biochem. J.* 133:529.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled EVL1, washed, and any wells with labeled EVL1 complex are assayed. Data obtained

using different concentrations of EVL1 are used to calculate values for the number, affinity, and association of EVL1 with the candidate molecules.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 4

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 418 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: HEAONOT03
- (B) CLONE: 3089412

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

Met Ala Thr Ser Glu Gln Ser Ile Cys Gln Ala Arg Ala Ser Val Met
 1           5           10           15

Val Tyr Asp Asp Thr Ser Lys Lys Trp Val Pro Ile Lys Pro Gly Gln
20           25           30

Gln Gly Phe Ser Arg Ile Asn Ile Tyr His Asn Thr Ala Ser Asn Thr
35           40           45

Phe Arg Val Val Gly Val Lys Leu Gln Asp Gln Gln Val Val Ile Asn
50           55           60

Tyr Ser Ile Val Lys Gly Leu Lys Tyr Asn Gln Ala Thr Pro Thr Phe
65           70           75           80

His Gln Trp Arg Asp Ala Arg Gln Val Tyr Gly Leu Asn Phe Ala Ser
85           90           95

Lys Glu Glu Ala Thr Thr Phe Ser Asn Ala Met Leu Phe Ala Leu Asn
100          105          110

Ile Met Asn Ser Gln Glu Gly Gly Pro Ser Ser Gln Arg Gln Val Gln
115          120          125

Asn Gly Pro Ser Pro Asp Glu Met Asp Ile Gln Arg Arg Gln Val Met
130          135          140

Glu Gln His Gln Gln Arg Gln Glu Ser Leu Glu Arg Arg Thr Ser
145          150          155          160

Ala Thr Gly Pro Ile Leu Pro Pro Gly His Pro Ser Ser Ala Ala Ser
165          170          175

Ala Pro Val Ser Cys Ser Gly Pro Pro Pro Pro Pro Pro Leu Val
180          185          190

Pro Pro Pro Pro Thr Gly Ala Thr Pro Pro Pro Pro Pro Leu Pro

```

-continued

195	200	205
Ala Gly Gly Ala Gln Gly Ser Ser His Asp Glu Ser Ser Met Ser Gly		
210	215	220
Leu Ala Ala Ala Ile Ala Gly Ala Lys Leu Arg Arg Val Gln Arg Pro		
225	230	235
Glu Asp Ala Ser Gly Gly Ser Ser Pro Ser Gly Thr Ser Lys Ser Asp		
	245	250
Ala Asn Arg Ala Ser Ser Gly Gly Gly Gly Gly Gly Leu Met Glu Glu		
	260	270
Met Asn Lys Leu Leu Ala Lys Arg Arg Lys Ala Ala Ser Gln Ser Asp		
	275	280
Lys Pro Ala Glu Lys Lys Glu Asp Glu Ser Gln Met Glu Asp Pro Ser		
	290	295
Thr Ser Pro Ser Pro Gly Thr Arg Ala Ala Ser Gln Pro Pro Asn Ser		
	305	310
Ser Glu Ala Gly Arg Lys Pro Trp Glu Arg Ser Asn Ser Val Glu Lys		
	315	320
Pro Val Ser Ser Ile Leu Ser Arg Thr Pro Ser Val Ala Lys Ser Pro		
	325	330
Glu Ala Lys Ser Pro Leu Gln Ser Gln Pro His Ser Arg Met Lys Pro		
	335	340
Ala Gly Ser Val Asn Asp Met Ala Leu Asp Ala Phe Asp Leu Asp Arg		
	345	350
Met Lys Gln Glu Ile Leu Glu Glu Val Val Arg Glu Leu His Lys Val		
	355	360
Lys Glu Glu Ile Ile Asp Ala Ile Arg Gln Glu Leu Ser Gly Ile Ser		
	365	370
	375	380
	385	390
	395	400
	405	410
		415
Thr Thr		

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1889 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:
- (A) LIBRARY: HEAONOT03
 - (B) CLONE: 3089412

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TTTAAGTAGG CTATAAAAT CAAGTTGCTG TCTTCAGAGG GTCTGTGGTC CTCTGATCAA	60
CATAGGCTGG TGGGAGTACA GGAATCGCCT CCTCAGGGTT CCCTGTGCTG CCACTTTTCA	120
GCCATGGCCA CAAGTGAACA GAGTATCTGC CAAGCCCGGG CTTCCTGTGAT GGTCTACGAT	180
GACACCAGTA AGAAATGGGT ACCAATCAAA CCTGGCCAGC AGGGATTGAG CCGGATCAAC	240
ATCTACCACA AACTGCCCAG CAACACCTTC AGAGTCGTTG GAGTCAAGTT GCAGGATCAG	300
CAGGTTGTGA TCAATTATTC AATCGTGAAA GGGCTGAAAT ACAATCAGGC CACGCCAACC	360
TTCCACCAGT GGCAGATGC CCGCCAGGTC TACGGCTTAA ACTTTGCAAG TAAAGAAGAG	420
GCAACCACGT TCTCCAATGC AATGCTGTTT GCCCTGAACA TCATGAATTC CCAAGAAGGA	480
GGCCCTCCA GCCAGCGTCA GGTGCAGAA GTGCCCTCTC CTGATGAGAT GGACATCCAG	540
AGAAGACAAG TGATGGAGCA GCACCAGCAG CAGCGTCAGG AATCTCTAGA AAGAAGAACC	600
TCGGCCACAG GGCCATCCT CCCACCAGGA CATCTTCAT CTGCAGCCAG CGCCCCCGTC	660

-continued

```

TCATGTAGTG GGCCTCCACC GCCCCCCCCA CCTCTAGTCC CACCTCCACC CACTGGGGCT 720
ACCCACACCTC CCCCACCCCC ACTGCCAGCC GGAGGAGCCC AGGGGTCCAG CCACGACGAG 780
AGCTCCATGT CAGGACTGGC CGCTGCCATA GCTGGGGCCA AGCTGAGAAG AGTCCAACGG 840
CCAGAAGACG CATCTGGAGG CTCCTAGTCCC AGTGGGACCT CAAAGTCCGA TGCCAACCGG 900
GCAAGCAGCG GGGGTGGCGG AGGAGGCCTC ATGGAGGAAA TGAACAACT GCTGGCCAAG 960
AGGAGAAAAG CAGCCTCCCA GTCAGACAAG CCAGCCGAGA AGAAGGAAGA TGAAAGCCAA 1020
ATGGAAGATC CTAGTACCTC CCCCTCTCCG GGGACCCGAG CAGCCAGCCA GCCACCTAAC 1080
TCCTCAGAGG CTGGCCGGAA GCCCTGGGAG CGGAGCAACT CGGTGGAGAA GCCTGTGTCC 1140
TCGATTCTGT CCAGAACCCC GTCTGTGGCA AAGAGCCCCG AAGCTAAGAG CCCCCTTCAG 1200
TCGCAGCCTC ACTCTAGGAT GAAGCCTGCT GGGAGCGTGA ATGACATGGC CCTGGATGCC 1260
TTGCACTTGG ACCGGATGAA GCAGGAGATC CTAGAGGAGG TGGTGAGAGA GCTCCACAAG 1320
GTGAAGGAGG AGATCATCGA CGCCATCAGG CAGGAGCTGA GTGGGATCAG CACCACGTAA 1380
GGGGCCGCGC TCCTCGCGCT GATTCGTCGA GCCCATCCGG CGACAGAGGA CAGCCAGAAG 1440
CCCAGCCAGC CCCAGACTCC AGTGCACCAG AGCACGCACA GGAGCCTGGG CGCGCTGCTG 1500
TGAAACGTCC TGACCTGTGA TCACACATGA CAGTGAGGAA ACCAAGTGCA ACTCCTGGGT 1560
TTTTTTTAGA TTCTGCCTGA CACGGAACAC CAGGTCTGCT CGTCTTTTTT GTGTTTTATA 1620
TTTGCTTATT TAAGGTACAT TTCTTTGGGT TTCTAGAGAC GCCCCTAAGT CACCTGCTTC 1680
ATTAGACGGT TTCCAGGTTT TCTCCAGGT GACGCTGTTA GCGCCTCAGC TGGCGGTGAC 1740
AGCCGCCCCA GCGTGCGGCC ACCACACACC GCAGAGCTGT CCAGGCACAG CTCCTCCCCC 1800
AGCGCTCATG GTGTTGAAAC TGTCTGTCAT GCACCACGGT GTCTGTGTCC ACACAGTAAT 1860
AAACGGTTTA CTGTCCGCAA AAAAAAAAAA 1889

```

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 393 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

- (vii) IMMEDIATE SOURCE:
 (A) LIBRARY: GenBank
 (B) CLONE: 1644453

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

Met Ser Glu Gln Ser Ile Cys Gln Ala Arg Ala Ser Val Met Val Tyr
 1           5           10          15
Asp Asp Thr Ser Lys Lys Trp Val Pro Ile Lys Pro Gly Gln Gln Gly
 20          25          30
Phe Ser Arg Ile Asn Ile Tyr His Asn Thr Ala Ser Ser Thr Phe Arg
 35          40          45
Val Val Gly Val Lys Leu Gln Asp Gln Gln Val Val Ile Asn Tyr Ser
 50          55          60
Ile Val Lys Gly Leu Lys Tyr Asn Gln Ala Thr Pro Thr Phe His Gln
 65          70          75          80
Trp Arg Asp Ala Arg Gln Val Tyr Gly Leu Asn Phe Ala Ser Lys Glu
 85          90          95
Glu Ala Thr Thr Phe Ser Asn Ala Met Leu Phe Ala Leu Asn Ile Met
100          105          110

```

-continued

```

Asn Ser Gln Glu Gly Gly Pro Ser Thr Gln Arg Gln Val Gln Asn Gly
 115                      120                      125

Pro Ser Pro Glu Glu Met Asp Ile Gln Arg Arg Gln Val Met Glu Gln
 130                      135                      140

Gln His Arg Gln Glu Ser Leu Glu Arg Arg Ile Ser Ala Thr Gly Pro
 145                      150                      155                      160

Ile Leu Pro Pro Gly His Pro Ser Ser Ala Ala Ser Thr Thr Leu Ser
      165                      170                      175

Cys Ser Gly Pro Pro Pro Pro Pro Pro Pro Pro Val Pro Pro Pro Pro
      180                      185                      190

Thr Gly Ser Thr Pro Pro Pro Pro Pro Pro Pro Leu Pro Ala Gly Gly Ala
      195                      200                      205

Gln Gly Thr Asn His Asp Glu Ser Ser Ala Ser Gly Leu Ala Ala Ala
 210                      215                      220

Leu Ala Gly Ala Lys Leu Arg Arg Val Gln Arg Pro Glu Asp Ala Ser
 225                      230                      235                      240

Gly Gly Ser Ser Pro Ser Gly Thr Ser Lys Ser Asp Ala Asn Arg Ala
      245                      250                      255

Ser Ser Gly Gly Gly Gly Gly Gly Leu Met Glu Glu Met Asn Lys Leu
      260                      265                      270

Leu Ala Lys Arg Arg Lys Ala Ala Ser Gln Thr Asp Lys Pro Ala Asp
      275                      280                      285

Arg Lys Glu Asp Glu Ser Gln Thr Glu Asp Pro Ser Thr Ser Pro Ser
      290                      295                      300

Pro Gly Thr Arg Ala Thr Ser Gln Pro Pro Asn Ser Ser Glu Ala Gly
 305                      310                      315                      320

Arg Lys Pro Trp Glu Arg Ser Asn Ser Val Glu Lys Pro Val Ser Ser
      325                      330                      335

Leu Leu Ser Arg Val Lys Pro Ala Gly Ser Val Asn Asp Val Gly Leu
      340                      345                      350

Asp Ala Leu Asp Leu Asp Arg Met Lys Gln Glu Ile Leu Glu Glu Val
      355                      360                      365

Val Arg Glu Leu His Lys Val Lys Glu Glu Ile Ile Asp Ala Ile Arg
      370                      375                      380

Gln Glu Leu Ser Gly Ile Ser Thr Thr
 385                      390

```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 380 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:

- (A) LIBRARY: GenBank
- (B) CLONE: 624964

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

Met Ser Glu Thr Val Ile Cys Ser Ser Arg Ala Thr Val Met Leu Tyr
 1                      5                      10                      15

Asp Asp Gly Asn Lys Arg Trp Leu Pro Ala Gly Thr Gly Pro Gln Ala
      20                      25                      30

Phe Ser Arg Val Gln Ile Tyr His Asn Pro Thr Ala Asn Ser Phe Arg
      35                      40                      45

Val Val Gly Arg Lys Met Gln Pro Asp Gln Gln Val Val Ile Asn Cys

```

-continued

50	55	60
Ala Ile Val Arg Gly Val Lys Tyr Asn Gln Ala Thr Pro Asn Phe His		
65	70	75 80
Gln Trp Arg Asp Ala Arg Gln Val Trp Gly Leu Asn Phe Gly Ser Lys		
	85	90 95
Glu Asp Ala Ala Gln Phe Ala Ala Gly Met Ala Ser Ala Leu Glu Ala		
	100	105 110
Leu Glu Gly Gly Gly Pro Pro Pro Pro Ala Leu Pro Thr Trp Ser		
	115	120 125
Val Pro Asn Gly Pro Ser Pro Glu Glu Val Glu Gln Gln Lys Arg Gln		
	130	135 140
Gln Pro Gly Pro Ser Glu His Ile Glu Arg Arg Val Ser Asn Ala Gly		
	145	150 155 160
Gly Pro Pro Ala Pro Pro Ala Gly Gly Pro Pro Pro Pro Gly Pro		
	165	170 175
Pro Pro Pro Pro Gly Pro Pro Pro Pro Gly Leu Pro Pro Ser Gly		
	180	185 190
Val Pro Ala Ala Ala His Gly Ala Gly Gly Gly Pro Pro Pro Ala Pro		
	195	200 205
Pro Leu Pro Ala Ala Gln Gly Pro Gly Gly Gly Gly Ala Gly Ala Pro		
	210	215 220
Gly Leu Ala Ala Ala Ile Ala Gly Ala Lys Leu Arg Lys Val Ser Lys		
	225	230 235 240
Gln Glu Glu Ala Ser Gly Gly Pro Thr Ala Pro Lys Ala Glu Ser Gly		
	245	250 255
Arg Ser Gly Gly Gly Gly Leu Met Glu Glu Met Asn Ala Met Leu Ala		
	260	265 270
Arg Arg Arg Lys Ala Thr Gln Val Gly Glu Lys Thr Pro Lys Asp Glu		
	275	280 285
Ser Ala Asn Gln Glu Glu Pro Glu Ala Arg Val Pro Ala Gln Ser Glu		
	290	295 300
Ser Val Arg Arg Pro Trp Glu Lys Asn Ser Thr Thr Leu Pro Arg Met		
	305	310 315 320
Lys Ser Ser Ser Ser Val Thr Thr Ser Glu Thr Gln Pro Cys Thr Pro		
	325	330 335
Ser Ser Ser Asp Tyr Ser Asp Leu Gln Arg Val Lys Gln Glu Leu Leu		
	340	345 350
Glu Glu Val Lys Lys Glu Leu Gln Lys Val Lys Glu Glu Ile Ile Glu		
	355	360 365
Ala Phe Val Gln Glu Leu Arg Lys Arg Gly Ser Pro		
	370	375 380

What is claimed is:

1. A substantially purified polypeptide comprising an amino acid sequence selected from the group consisting of
 - a) an amino acid sequence of SEQ ID NO:1,
 - b) a naturally-occurring amino acid sequence having at least 95% sequence identity to the sequence of SEQ ID

55

- NO:1, wherein said amino acid sequence encodes a polypeptide that binds the neural protein FE65.
2. A composition comprising a polypeptide of claim 1 in conjunction with a suitable pharmaceutical carrier.

* * * * *



US006498235B2

(12) **United States Patent**
Sheppard et al.

(10) **Patent No.:** **US 6,498,235 B2**
(45) **Date of Patent:** **Dec. 24, 2002**

(54) **TESTIS SPECIFIC GLYCOPROTEIN ZPEP10**

(75) **Inventors:** **Paul O. Sheppard**, Redmond, WA (US); **Christopher S. Piddington**, Thousand Oaks, CA (US); **Jeff L. Ellsworth**, Seattle, WA (US)

(73) **Assignee:** **ZymoGenetics, Inc.**, Seattle, WA (US)

(*) **Notice:** Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 40 days.

(21) **Appl. No.:** **09/789,453**

(22) **Filed:** **Feb. 20, 2001**

(65) **Prior Publication Data**

US 2002/0102704 A1 Aug. 1, 2002

Related U.S. Application Data

(62) Division of application No. 09/441,346, filed on Nov. 16, 1999, now Pat. No. 6,242,588.

(60) Provisional application No. 60/109,216, filed on Nov. 20, 1998.

(51) **Int. Cl.⁷** **C07K 1/00; C07K 14/00; C07K 17/08; C07K 2/00; A61K 38/00**

(52) **U.S. Cl.** **530/350; 530/300; 530/324; 514/1; 514/2; 514/4**

(58) **Field of Search** 530/300, 324, 530/350; 514/1, 2, 4

(56) **References Cited**
PUBLICATIONS

U.S. patent application Ser. No. 09/441,346, Sheppard et al., filed Nov. 16, 1999.

Wilson, WashU-Merck EST Project, 1997, GenBank ACC#AA459848.

Strausberg, NCI, Cancer Genome Anatomy Project, 1998, GenBank Acc#AA868533.

Lexicon Pharmaceuticals, Mouse OST, 1998, OST16697.

Lexicon Pharmaceuticals, Mouse OST, 1998, OST16698.

Primary Examiner—Yvonne Eyler

Assistant Examiner—Alana M. Harris

(74) *Attorney, Agent, or Firm*—Brian J. Walsh

(57) **ABSTRACT**

The present invention relates to zpep10 polypeptides and polynucleotides encoding the same. Zpep10 polypeptide is a testis-specific membrane glycoprotein. Zpep10 polypeptides would be useful for modulating spermatogenesis and egg-sperm interaction and would be useful to study or modulate these functions in in vitro or in vivo systems. The present invention also includes antibodies to the zpep10 polypeptides.

13 Claims, 3 Drawing Sheets

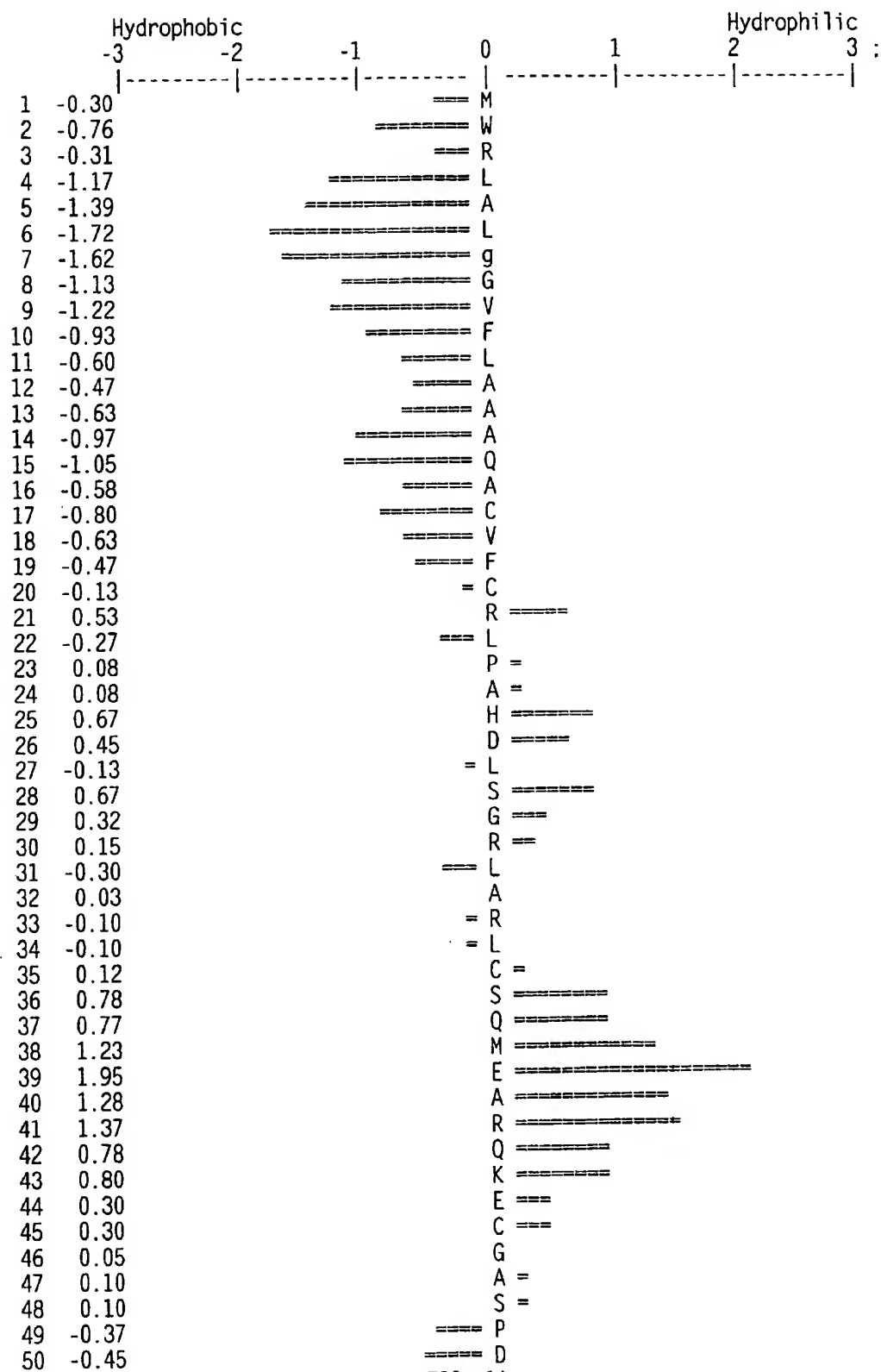
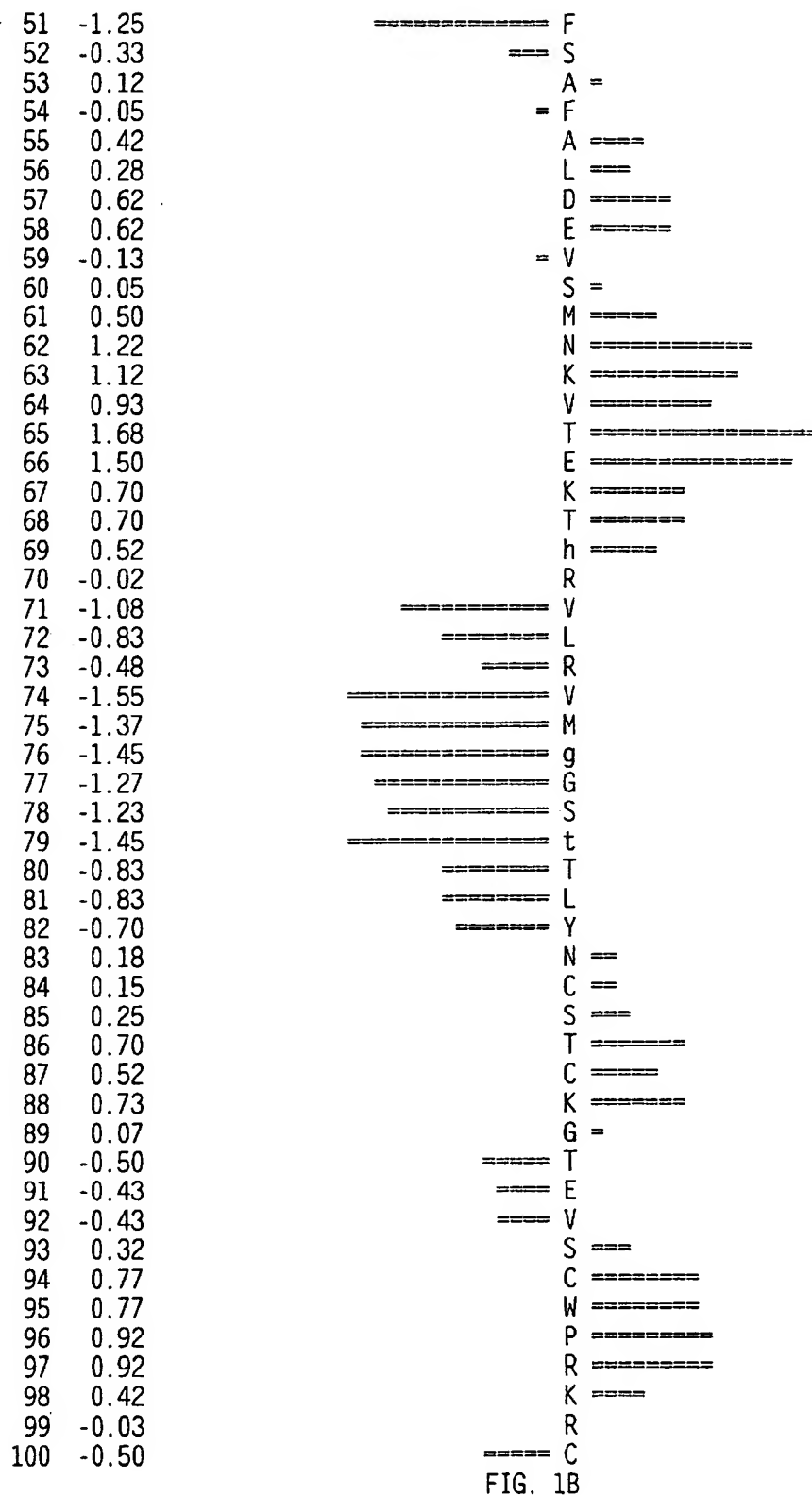


FIG. 1A



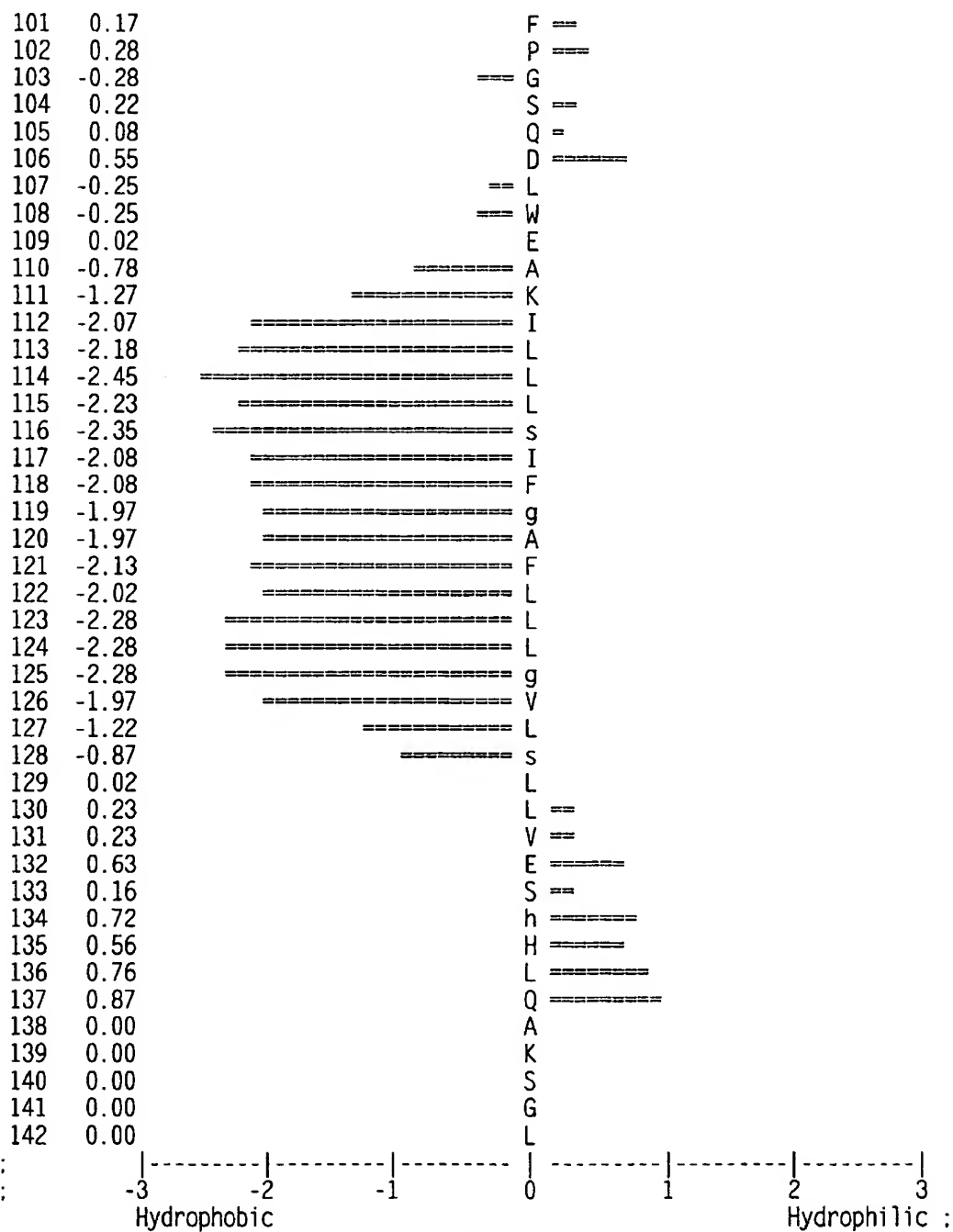


Fig. 1C

TESTIS SPECIFIC GLYCOPROTEIN ZPEP10

REFERENCE TO RELATED APPLICATIONS

This application is a divisional of application Ser. No. 09/441,346 filed Nov. 16, 1999 now issued as U.S. Pat. No. 6,242,588 which is related to Provisional Application No. 60/109,216, filed on Nov. 20, 1998. Under 35 U.S.C. § 119(e)(1), this application claims benefit of said Provisional Application.

BACKGROUND OF THE INVENTION

The testis is the center for spermatogenesis, the process by which a germ cell proceeds through multiple stages of differentiation, and culminates in the formation of a terminally differentiated cell (spermatozoa or sperm) having a unique function. Within the testis are seminiferous tubules, where spermatogonium mature into spermatozoa. Surrounding the seminiferous tubules are interstitial cells which secrete androgens, such as testosterone, required for maturation and function of the testis and development of secondary sexual characteristics. Disorders of the testis are common and have profound effect. Infertility can result from disorders occurring during spermatogenesis. Many developmental disorders, such as hypogonadism, are associated with altered sex hormone production and levels in the testis. Testicular cancer, although rare, is the most common form of cancer in young men between the ages of 15 and 35.

Testis specific proteins have therapeutic value in the treatment of disorders associated with the testis such as dysfunctional sperm production, infertility and testicular cancer. Towards this end, the present invention provides novel testis-specific membrane glycoproteins, soluble ligands, agonists and antagonists, related compositions and methods as well as other uses that should be apparent to those skilled in the art from the teachings herein.

BRIEF DESCRIPTION OF THE DRAWING

FIGS. 1A-K are a Hopp/Woods hydrophilicity profile of the amino acid sequence shown in SEQ ID NO:2. The profile is based on a sliding six-residue window. Buried G, S, and T residues and exposed H, Y, and W residues were ignored. These residues are indicated in the FIG. by lower case letters.

DETAILED DESCRIPTION OF THE INVENTION

Prior to setting forth the invention in detail, it may be helpful to the understanding thereof to define the following terms:

The term "affinity tag" is used herein to denote a peptide segment that can be attached to a polypeptide to provide for purification or detection of the polypeptide or provide sites for attachment of the polypeptide to a substrate. In principal, any peptide or protein for which an antibody or other specific binding agent is available can be used as an affinity tag. Affinity tags include a poly-histidine tract, protein A (Nilsson et al., *EMBO J.* 4:1075, 1985; Nilsson et al., *Methods Enzymol.* 198:3, 1991), glutathione S transferase (Smith and Johnson, *Gene* 67:31, 1988), Glu—Glu affinity tag (Grussenmeyer et al., *Proc. Natl. Acad. Sci. USA* 82:7952-4, 1985), substance P, Flag™ peptide (Hopp et al., *Biotechnology* 6:1204-10, 1988; available from Eastman Kodak Co., New Haven, Conn.), streptavidin binding peptide, or other antigenic epitope or binding domain. See, in general Ford et al., *Protein Expression and Purification* 2:

95-107, 1991. DNAs encoding affinity tags are available from commercial suppliers (e.g., Pharmacia Biotech, Piscataway, N.J.).

The term "allelic variant" denotes any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequence. The term allelic variant is also used herein to denote a protein encoded by an allelic variant of a gene.

The terms "amino-terminal" and "carboxyl-terminal" are used herein to denote positions within polypeptides and proteins. Where the context allows, these terms are used with reference to a particular sequence or portion of a polypeptide or protein to denote proximity or relative position. For example, a certain sequence positioned carboxyl-terminal to a reference sequence within a protein is located proximal to the carboxyl terminus of the reference sequence, but is not necessarily at the carboxyl terminus of the complete protein.

The term "complements of polynucleotide molecules" denotes polynucleotide molecules having a complementary base sequence and reverse orientation as compared to a reference sequence. For example, the sequence 5' ATG-CACGGG 3' is complementary to 5' CCCGTGCAT 3'.

The term "contig" denotes a polynucleotide that has a contiguous stretch of identical or complementary sequence to another polynucleotide. Contiguous sequences are said to "overlap" a given stretch of polynucleotide sequence either in their entirety or along a partial stretch of the polynucleotide. For example, representative contigs to the polynucleotide sequence 5'-ATGGCTTAGCTT-3' (SEQ ID NO:12) are 5'-TAGCTTgagctt-3' (SEQ ID NO: 13) and 3'-gtcgacTACCGA-5' (SEQ ID NO: 14).

The term "degenerate nucleotide sequence" denotes a sequence of nucleotides that includes one or more degenerate codons (as compared to a reference polynucleotide molecule that encodes a polypeptide) Degenerate codons contain different triplets of nucleotides, but encode the same amino acid residue (i.e., GAU and GAC triplets each encode Asp).

The term "expression vector" denotes a DNA molecule, linear or circular, that comprises a segment encoding a polypeptide of interest operably linked to additional segments that provide for its transcription. Such additional segments may include promoter and terminator sequences, and may optionally include one or more origins of replication, one or more selectable markers, an enhancer, a polyadenylation signal, and the like. Expression vectors are generally derived from plasmid or viral DNA, or may contain elements of both.

The term "isolated", when applied to a polynucleotide, denotes that the polynucleotide has been removed from its natural genetic milieu and is thus free of other extraneous or unwanted coding sequences, and is in a form suitable for use within genetically engineered protein production systems. Such isolated molecules are those that are separated from their natural environment and include cDNA and genomic clones. Isolated DNA molecules of the present invention are free of other genes with which they are ordinarily associated, but may include naturally occurring 5' and 3' untranslated regions such as promoters and terminators. The identification of associated regions will be evident to one of ordinary skill in the art (see for example, Dynan and Tijan, *Nature* 316:774-78, 1985).

An "isolated" polypeptide or protein is a polypeptide or protein that is found in a condition other than its native environment, such as apart from blood and animal tissue. In a preferred form, the isolated polypeptide is substantially free of other polypeptides, particularly other polypeptides of animal origin. It is preferred to provide the polypeptides in a highly purified form, i.e. greater than 95% pure, more preferably greater than 99% pure. When used in this context, the term "isolated" does not exclude the presence of the same polypeptide in alternative physical forms, such as dimers or alternatively glycosylated or derivatized forms.

The term "operably linked", when referring to DNA segments, denotes that the segments are arranged so that they function in concert for their intended purposes, e.g. transcription initiates in the promoter and proceeds through the coding segment to the terminator.

The term "ortholog" denotes a polypeptide or protein obtained from one species that is the functional counterpart of a polypeptide or protein from a different species. Sequence differences among orthologs are the result of speciation.

The term "polynucleotide" denotes a single- or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. Polynucleotides include RNA and DNA, and may be isolated from natural sources, synthesized *in vitro*, or prepared from a combination of natural and synthetic molecules. Sizes of polynucleotides are expressed as base pairs (abbreviated "bp"), nucleotides ("nt"), or kilobases ("kb"). Where the context allows, the latter two terms may describe polynucleotides that are single-stranded or double-stranded. When the term is applied to double-stranded molecules it is used to denote overall length and will be understood to be equivalent to the term "base pairs". It will be recognized by those skilled in the art that the two strands of a double-stranded polynucleotide may differ slightly in length and that the ends thereof may be staggered as a result of enzymatic cleavage; thus all nucleotides within a double-stranded polynucleotide molecule may not be paired. Such unpaired ends will in general not exceed 20 nt in length.

A "polypeptide" is a polymer of amino acid residues joined by peptide bonds, whether produced naturally or synthetically. Polypeptides of less than about 10 amino acid residues are commonly referred to as "peptides".

"Probes and/or primers" as used herein can be RNA or DNA. DNA can be either cDNA or genomic DNA. Polynucleotide probes and primers are single or double-stranded DNA or RNA, generally synthetic oligonucleotides, but may be generated from cloned cDNA or genomic sequences or its complements. Analytical probes will generally be at least 20 nucleotides in length, although somewhat shorter probes (14-17 nucleotides) can be used. PCR primers are at least 5 nucleotides in length, preferably 15 or more nt, more preferably 20-30 nt. Short polynucleotides can be used when a small region of the gene is targeted for analysis. For gross analysis of genes, a polynucleotide probe may comprise an entire exon or more. Probes can be labeled to provide a detectable signal, such as with an enzyme, biotin, a radionuclide, fluorophore, chemiluminescer, paramagnetic particle and the like, which are commercially available from many sources, such as Molecular Probes, Inc., Eugene, Oreg., and Amersham Corp., Arlington Heights, Ill., using techniques that are well known in the art.

The term "promoter" denotes a portion of a gene containing DNA sequences that provide for the binding of RNA polymerase and initiation of transcription. Promoter

sequences are commonly, but not always, found in the 5' non-coding regions of genes.

A "protein" is a macromolecule comprising one or more polypeptide chains. A protein may also comprise non-peptidic components, such as carbohydrate groups. Carbohydrates and other non-peptidic substituents may be added to a protein by the cell in which the protein is produced, and will vary with the type of cell. Proteins are defined herein in terms of their amino acid backbone structures; substituents such as carbohydrate groups are generally not specified, but may be present nonetheless.

The term "receptor" denotes a cell-associated protein that binds to a bioactive molecule (i.e., a ligand) and mediates the effect of the ligand on the cell. Membrane-bound receptors are characterized by a multi-domain structure comprising an extracellular ligand-binding domain and an intracellular effector domain that is typically involved in signal transduction. Binding of ligand to receptor results in a conformational change in the receptor that causes an interaction between the effector domain and other molecule(s) in the cell. This interaction in turn leads to an alteration in the metabolism of the cell. Metabolic events that are linked to receptor-ligand interactions include gene transcription, phosphorylation, dephosphorylation, increases in cyclic AMP production, mobilization of cellular calcium, mobilization of membrane lipids, cell adhesion, hydrolysis of inositol lipids and hydrolysis of phospholipids. Most nuclear receptors also exhibit a multi-domain structure, including an amino-terminal, transactivating domain, a DNA binding domain and a ligand binding domain. In general, receptors can be membrane bound, cytosolic or nuclear; monomeric (e.g., thyroid stimulating hormone receptor, beta-adrenergic receptor) or multimeric (e.g., PDGF receptor, growth hormone receptor, IL-3 receptor, GM-CSF receptor, G-CSF receptor, erythropoietin receptor and IL-6 receptor).

The term "secretory signal sequence" denotes a DNA sequence that encodes a polypeptide (a "secretory peptide") that, as a component of a larger polypeptide, directs the larger polypeptide through a secretory pathway of a cell in which it is synthesized. The larger peptide is commonly cleaved to remove the secretory peptide during transit through the secretory pathway.

The term "soluble receptor" is used herein to refer to a receptor polypeptide that is not bound to a cell membrane. Soluble receptors are most commonly ligand-binding receptor polypeptides that lack transmembrane and cytoplasmic domains. Soluble receptors can comprise additional amino acid residues, such as affinity tags that provide for purification of the polypeptide or provide sites for attachment of the polypeptide to a substrate. Many cell-surface receptors have naturally occurring, soluble counterparts that are produced by proteolysis or translated from alternatively spliced mRNAs. Receptor polypeptides are said to be substantially free of transmembrane and intracellular polypeptide segments when they lack sufficient portions of these segments to provide membrane anchoring or signal transduction, respectively.

The term "splice variant" is used herein to denote alternative forms of RNA transcribed from a gene. Splice variation arises naturally through use of alternative splicing sites within a transcribed RNA molecule, or less commonly between separately transcribed RNA molecules, and may result in several mRNAs transcribed from the same gene. Splice variants may encode polypeptides having altered amino acid sequence. The term splice variant is also used herein to denote a protein encoded by a splice variant of an mRNA transcribed from a gene.

Molecular weights and lengths of polymers determined by imprecise analytical methods (e.g., gel electrophoresis) will be understood to be approximate values. When such a value is expressed as "about" X or "approximately" X, the stated value of X will be understood to be accurate to $\pm 10\%$.

All references cited herein are incorporated by reference in their entirety.

Within one aspect the invention provides an isolated polypeptide comprising an extracellular domain, wherein the extracellular domain comprises amino acid residues 22 to 111 of the amino acid sequence of SEQ ID NO:2. Within one embodiment polypeptide further comprises a transmembrane domain that resides in a carboxyl-terminal position relative to the extracellular domain, wherein the transmembrane domain comprises amino acid residues 112 to 133 of the amino acid sequence of SEQ ID NO:2. Within another embodiment the polypeptide further comprises a cytoplasmic domain that resides in a carboxyl-terminal position relative to the transmembrane domain, wherein the cytoplasmic domain comprises amino acid residues 134 to 142 of the amino acid sequence of SEQ ID NO:2. Within another embodiment the polypeptide further comprises a secretory signal that resides in an amino-terminal position relative to the extracellular domain, wherein the secretory signal sequence comprises amino acid residues 1 to 20 of the amino acid sequence of SEQ ID NO:2.

The invention also provides an isolated polypeptide as described herein comprising amino acid residue 1 to amino acid residue 142 of SEQ ID NO:2.

Also provided is an isolated polypeptide as described herein, covalently linked amino terminally or carboxy terminally to a moiety selected from the group consisting of affinity tags, toxins, radionucleotides, enzymes and fluorophores.

Within another aspect the invention provides an isolated polypeptide comprising a sequence of amino acid residues that is at least 80% identical to a amino acid residue 21 to amino acid residue 142 of SEQ ID NO:2, wherein the polypeptide specifically binds with an antibody that specifically binds with a polypeptide having the amino acid sequence of SEQ ID NO:2. Within one embodiment any difference between the amino acid sequence of the isolated polypeptide and the corresponding amino acid sequence of SEQ ID NO:2 is due to a conservative amino acid substitution. Within another embodiment the amino acid percent identity is determined using a FASTA program with ktup=1, gap opening penalty=10, gap extension penalty=1, and substitution matrix=blosum62, with other parameters set as default.

The invention provides an isolated polypeptide comprising the amino acid sequence of amino acid residue 1 to amino acid residue 20 of SEQ ID NO:2.

Also provided is an isolated polypeptide selected from the group consisting of:

- a) amino acid residues 21–111 of SEQ ID NO:2;
- b) amino acid residues 112–133 of SEQ ID NO:2;
- c) amino acid residues 134–142 of SEQ ID NO:2;
- d) amino acid residues 1–20 of SEQ ID NO:2;
- e) amino acid residues 21–133 of SEQ ID NO:2;
- f) amino acid residues 112–142 of SEQ ID NO:2;
- g) amino acid residues 1–111 of SEQ ID NO:2; and
- h) amino acid residues 1–133 of SEQ ID NO:2.

Within another aspect the invention provides a fusion protein consisting of a first portion and a second portion

joined by a peptide bond, the first portion comprising a polypeptide as described herein, and the second portion comprising another polypeptide.

The invention also provides a polypeptide as described herein in combination with a pharmaceutically acceptable vehicle.

Within another aspect the invention provides an antibody that specifically binds to an epitope of a polypeptide of as described herein. Within one embodiment the antibody is selected from the group consisting of: a) polyclonal antibody; b) murine monoclonal antibody; c) humanized antibody derived from b); and d) human monoclonal antibody.

Within another embodiment the antibody fragment is selected from the group consisting of F(ab'), F(ab), Fab', Fab, Fv, scFv, and minimal recognition unit.

Also provided is an anti-idiotypic antibody that specifically binds to an antibody as described herein. Also provided is a binding protein that specifically binds to an epitope of a polypeptide as described herein.

Within another aspect the invention provides a method of producing an antibody to a polypeptide comprising: inoculating an animal with a polypeptide as described herein; wherein the polypeptide elicits an immune response in the animal to produce the antibody; and isolating the antibody from the animal.

Within another aspect is provided an isolated polynucleotide encoding a polypeptide comprising an extracellular domain, wherein the extracellular domain comprises amino acid residues 22 to 111 of the amino acid sequence of SEQ ID NO:2. Within one embodiment the polypeptide further comprises a transmembrane domain that resides in a carboxyl-terminal position relative to the extracellular domain, wherein the transmembrane domain comprises amino acid residues 112 to 133 of the amino acid sequence of SEQ ID NO:2. Within another embodiment the polypeptide further comprises a cytoplasmic domain that resides in a carboxyl-terminal position relative to the transmembrane domain, wherein the cytoplasmic domain comprises amino acid residues 134 to 142 of the amino acid sequence of SEQ ID NO:2. Within yet another embodiment the polypeptide further comprises a secretory signal that resides in an amino-terminal position relative to the extracellular domain, wherein the secretory signal sequence comprises amino acid residues 1 to 20 of the amino acid sequence of SEQ ID NO:2.

The invention also provides an isolated polynucleotide as described herein encoding a polypeptide comprising amino acid residue 1 to amino acid residue 142 of SEQ ID NO:2.

Also provided is an isolated polynucleotide as described herein, wherein the polypeptide is covalently linked amino terminally or carboxy terminally to a moiety selected from the group consisting of affinity tags, toxins, radionucleotides, enzymes and fluorophores.

Within another aspect the invention provides an isolated polynucleotide encoding a polypeptide comprising a sequence of amino acid residues that is at least 80% identical to a amino acid residue 21 to amino acid residue 142 of SEQ ID NO:2, wherein the polypeptide specifically binds with an antibody that specifically binds with a polypeptide having the amino acid sequence of SEQ ID NO:2. Within one embodiment any difference between the amino acid sequence of the isolated polypeptide and the corresponding amino acid sequence of SEQ ID NO:2 is due to a conservative amino acid substitution. Within another embodiment the amino acid percent identity is determined using a FASTA program with ktup=1, gap opening penalty=10, gap extension penalty=1, and substitution matrix=blosum62, with other parameters set as default.

The invention also provides an isolated polynucleotide selected from the group consisting of:

- a) a sequence of nucleotides from nucleotide 139 to nucleotide 411 of SEQ ID NO:1;
- b) a sequence of nucleotides from nucleotide 139 to nucleotide 477 of SEQ ID NO:1;
- c) a sequence of nucleotides from nucleotide 139 to nucleotide 504 of SEQ ID NO:1;
- d) a sequence of nucleotides from nucleotide 79 to nucleotide 504 of SEQ ID NO:1;
- e) a sequence of nucleotides from nucleotide 1 to nucleotide 1094 of SEQ ID NO:1;
- f) a polynucleotide that remains hybridized following stringent wash conditions to a polynucleotide consisting of the nucleotide sequence of SEQ ID NO:1, or the complement of SEQ ID NO:1; and
- g) nucleotide sequences complementary to a), b), c), d), e), or f.

Further provided is an isolated polynucleotide encoding a fusion protein consisting of a first portion and a second portion joined by a peptide bond, the first portion comprises a polypeptide as described herein; and the second portion comprising another polypeptide.

Also provided is an isolated polynucleotide encoding a fusion protein comprising a secretory signal sequence having the amino acid sequence of amino acid residues 1-20 of SEQ ID NO:2, wherein the secretory signal sequence is operably linked to an additional polypeptide.

The invention also provides an isolated polynucleotide comprising the sequence of nucleotide 1 to nucleotide 426 of SEQ ID NO:3.

Within another aspect is provided an expression vector comprising the following operably linked elements:

- a transcription promoter; a DNA segment encoding a polypeptide as described herein; and a transcription terminator.

Within one embodiment the DNA segment encodes a polypeptide covalently linked amino terminally or carboxy terminally to an affinity tag. Within another embodiment the DNA segment further encodes a secretory signal sequence operably linked to the polypeptide. Within yet another embodiment the secretory signal sequence comprises residues 1 to 20 of SEQ ID NO:2.

The invention also provides a cultured cell into which has been introduced an expression vector as described herein; wherein the cell expresses the polypeptide encoded by the DNA segment.

The invention also provides a method of producing a polypeptide comprising: culturing a cell into which has been introduced an expression vector as described herein; whereby the cell expresses the polypeptide encoded by the DNA segment; and recovering the expressed polypeptide.

The present invention is based in part upon the discovery of a novel DNA sequence (SEQ ID NO:1) and the corresponding deduced polypeptide sequence (SEQ ID NO:2) which encode a testis-specific polypeptide designated zpep10. The novel zpep10 polypeptide-encoding polynucleotides of the present invention were initially identified by querying an EST database for polypeptides containing repetitive patterns and post-translational processing sites yielding potentially active peptides. The polypeptide corresponding to an EST meeting those search criteria was further analyzed and found to be a membrane glycoprotein. The EST sequence was from a testis cell library. Several clones considered likely to contain the entire coding region were used for sequencing and resulted in an incompletely spliced

message. A minimal nucleotide sequence having all potential introns spliced out was generated. The full length cDNA sequence was identified from a testis library and is disclosed in SEQ ID NO:1. The deduced amino acid sequence of this polynucleotide sequence is disclosed in SEQ ID NO:2. Analysis of the DNA encoding a zpep10 polypeptide (SEQ ID NO:1) revealed an open reading frame encoding 142 amino acids (SEQ ID NO:2) comprising a putative signal sequence (residues 1 to 20 of SEQ ID NO:2, nucleotides 79 to 138 of SEQ ID NO:1) and 122 amino acids of predicted mature sequence (residues 21 to 142 of SEQ ID NO:2, nucleotides 139 to 504 of SEQ ID NO:1) containing an extracellular domain (residues 21 to 111 of SEQ ID NO:2, nucleotides 139 to 411 of SEQ ID NO:1) containing six cysteine residues, amino acid residues 35, 45, 84, 87, 94 and 100 of SEQ ID NO:2, a tri-basic amino acid cleavage site, amino acid residues 97-99 of SEQ ID NO:2; potential N-linked glycosylation sites at amino acid residues 83 and 86 of SEQ ID NO:2; and potential O-glycosylation sites at amino acid residues 28, 36, 48, 52, 60, 65, 68, 78, 79, 80, 85, 86, 90, 93 and 104 of SEQ ID NO:2; a putative transmembrane domain (residues 112 to 133 of SEQ ID NO:2, nucleotides 412 to 477 of SEQ ID NO:1) and a cytoplasmic domain (residues 134 to 142 of SEQ ID NO:2, nucleotides 478 to 504 of SEQ ID NO:1). The overall structure of zpep10 is helical. Those skilled in the art will recognize that these domain boundaries are approximate, and are based on alignments with known proteins and predictions of protein folding. Zpep10 does not share significant homology with any known protein.

Many proteins and hormones are processed into their mature forms by highly-specific proteolytic enzymes, prohormone convertases, which carry out intracellular cleavage at the COOH-terminal side of dibasic sites within their substrate polypeptides. There are only a few dibasic amino acid combinations, including lys-lys, arg-arg, arg-lys and lys-arg. Zpep10 polypeptides may be processed into an active form through cleavage after lys (amino acid residue 98 of SEQ ID NO:2) or arg (amino acid residue 99 of SEQ ID NO:2) of the tribasic site arg-lys-arg (amino acid residues 97-99 of SEQ ID NO:2). Prohormone convertase PC4 exhibits highly specific testis expression (WIPO publication, WO98/50560) and may serve to cleave the zpep10 polypeptide.

The present invention therefore provides post-translationally modified polypeptides or polypeptide fragments having the amino acid sequence from amino acid residue 21 to amino acid residue 98 of SEQ ID NO:2 and the amino acid sequence from amino acid residue 21 to amino acid residue 99 of SEQ ID NO:2. Examples of post translational modifications include proteolytic cleavage, glycosylation and disulfide bonding.

Analysis of the tissue distribution of the mRNA corresponding to this novel DNA by Northern blot and Dot blot analysis suggest that zpep10 is a testis-specific protein having a transcript of about 1.5 kb.

The present invention further provides polynucleotide molecules, including DNA and RNA molecules, encoding zpep10 proteins. The polynucleotides of the present invention include the sense strand; the anti-sense strand; and the DNA as double-stranded, having both the sense and anti-sense strand annealed together by their respective hydrogen bonds. Representative DNA sequences encoding zpep10 proteins are set forth in SEQ ID NO:1. DNA sequences encoding other zpep10 proteins can be readily generated by those of ordinary skill in the art based on the genetic code. Counterpart RNA sequences can be generated by substitution of U for T.

Those skilled in the art will readily recognize that, in view of the degeneracy of the genetic code, considerable sequence variation is possible among these polynucleotide molecules. SEQ ID NO:3 is a degenerate DNA sequence that encompasses all DNAs that encode the zpep10 polypeptide of SEQ ID NO:2. Those skilled in the art will recognize that the degenerate sequence of SEQ ID NO:3 also provides all RNA sequences encoding SEQ ID NO:2 by substituting U for T. Thus, zpep10 polypeptide-encoding polynucleotides comprising nucleotide 1 to nucleotide 426 of SEQ ID NO:3 and their RNA equivalents are contemplated by the present invention. Table 1 sets forth the one-letter codes used within SEQ ID NO:3 to denote degenerate nucleotide positions. "Resolutions" are the nucleotides denoted by a code letter. "Complement" indicates the code for the complementary nucleotide(s). For example, the code Y denotes either C or T, and its complement R denotes A or G, A being complementary to T, and G being complementary to C.

TABLE 1

Nucleotide	Resolution	Complement	Nucleotide
A	A	T	T
C	C	G	G
G	G	C	C
T	T	A	A
R	A G	Y	C T
Y	C T	R	A G
M	A C	K	G T
K	G T	M	A C
S	C G	S	C G
W	A T	W	A T
H	A C T	D	A G T
B	C G T	V	A C G
V	A C G	B	C G T
D	A G T	H	A C T
N	A C G T	N	A C G T

The degenerate codons used in SEQ ID NO:3, encompassing all possible codons for a given amino acid, are set forth in Table 2.

TABLE 2

Amino Acid	One Letter Code	Codons	Degenerate Codon
Cys	C	TGC TGT	TGY
Ser	S	AGC AGT TCA TCC TCG TCT	WSN
Thr	T	ACA ACC ACG ACT	ACN
Pro	P	CCA CCC CCG CCT	CCN
Ala	A	GCA GCC GCG GCT	GCN
Gly	G	GGA GGC GGG GGT	GGN
Asn	N	AAC AAT	AAV
Asp	D	GAC GAT	GAY
Glu	E	GAA GAG	GAR
Gln	Q	CAA CAG	CAR
His	H	CAC CAT	CAY
Arg	R	AGA AGG CGA CGC CGG CGT	MGN
Lys	K	AAA AAG	AAR
Met	M	ATG	ATG
Ile	I	ATA ATC ATT	ATH
Leu	L	CTA CTC CTG CTT TTA TTG	YTN
Val	V	GTA GTC GTG GTT	GTN
Phe	F	TTC TTT	TTY
Tyr	Y	TAC TAT	TAY
Trp	W	TGG	TGG
Ter	-	TAA TAG TGA	TRR
Asn Asp	B		RAY
Glu Gln	Z		SAR
Any	X		NNN

One of ordinary skill in the art will appreciate that some ambiguity is introduced in determining a degenerate codon,

representative of all possible codons encoding each amino acid. For example, the degenerate codon for serine (WSN) can, in some circumstances, encode arginine (AGR), and the degenerate codon for arginine (MGN) can, in some circumstances, encode serine (AGY). A similar relationship exists between codons encoding phenylalanine and leucine. Thus, some polynucleotides encompassed by the degenerate sequence may encode variant amino acid sequences, but one of ordinary skill in the art can easily identify such variant sequences by reference to the amino acid sequence of SEQ ID NO:2. Variant sequences can be readily tested for functionality as described herein.

One of ordinary skill in the art will also appreciate that different species can exhibit "preferential codon usage." In general, see, Grantham, et al., *Nuc. Acids Res.* 8:1893-912, 1980; Haas, et al. *Curr. Biol.* 6:315-24, 1996; Wain-Hobson, et al., *Gene* 13:355-64, 1981; Grosjean and Fiers, *Gene* 18:199-209, 1982; Holm, *Nuc. Acids Res.* 14:3075-87, 1986; Ikemura, *J. Mol. Biol.* 158:573-97, 1982. As used herein, the term "preferential codon usage" or "preferential codons" is a term of art referring to protein translation codons that are most frequently used in cells of a certain species, thus favoring one or a few representatives of the possible codons encoding each amino acid (See Table 2). For example, the amino acid threonine (Thr) may be encoded by ACA, ACC, ACG, or ACT, but in mammalian cells ACC is the most commonly used codon; in other species, for example, insect cells, yeast, viruses or bacteria, different Thr codons may be preferential. Preferential codons for a particular species can be introduced into the polynucleotides of the present invention by a variety of methods known in the art. Introduction of preferential codon sequences into recombinant DNA can, for example, enhance production of the protein by making protein translation more efficient within a particular cell type or species. Therefore, the degenerate codon sequence disclosed in SEQ ID NO:3 serves as a template for optimizing expression of polynucleotides in various cell types and species commonly used in the art and disclosed herein. Sequences containing preferential codons can be tested and optimized for expression in various species, and tested for functionality as disclosed herein.

Within preferred embodiments of the invention the isolated polynucleotides will hybridize to similar sized regions of SEQ ID NO:1, other polynucleotide probes, primers, fragments and sequences recited herein or sequences complementary thereto. Polynucleotide hybridization is well known in the art and widely used for many applications, see for example, Sambrook et al., *Molecular Cloning: A Laboratory Manual, Second Edition*, Cold Spring Harbor, N.Y., 1989; Ausubel et al., eds., *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc., NY, 1987; Berger and Kimmel, eds., *Guide to Molecular Cloning Techniques, Methods in Enzymology*, volume 152, 1987 and Wetmur, *Crit. Rev. Biochem. Mol. Biol.* 26:227-59, 1990. Polynucleotide hybridization exploits the ability of single stranded complementary sequences to form a double helix hybrid. Such hybrids include DNA-DNA, RNA-RNA and DNA-RNA.

Hybridization will occur between sequences which contain some degree of complementarity. Hybrids can tolerate mismatched base pairs in the double helix, but the stability of the hybrid is influenced by the degree of mismatch. The T_m of the mismatched hybrid decreases by 1° C. for every 1-1.5% base pair mismatch. Varying the stringency of the hybridization conditions allows control over the degree of mismatch that will be present in the hybrid. The degree of

stringency increases as the hybridization temperature increases and the ionic strength of the hybridization buffer decreases. Stringent hybridization conditions encompass temperatures of about 5–25° C. below the thermal melting point (T_m) of the hybrid and a hybridization buffer having up to 1 M Na⁺. Higher degrees of stringency at lower temperatures can be achieved with the addition of formamide which reduces the T_m of the hybrid about 1° C. for each 1% formamide in the buffer solution. Generally, such stringent conditions include temperatures of 20–70° C. and a hybridization buffer containing up to 6× SSC and 0–50% formamide. A higher degree of stringency can be achieved at temperatures of from 40–70° C. with a hybridization buffer having up to 4× SSC and from 0–50% formamide. Highly stringent conditions typically encompass temperatures of 42–70° C. with a hybridization buffer having up to 1× SSC and 0–50% formamide. Different degrees of stringency can be used during hybridization and washing to achieve maximum specific binding to the target sequence. Typically, the washes following hybridization are performed at increasing degrees of stringency to remove non-hybridized polynucleotide probes from hybridized complexes.

The above conditions are meant to serve as a guide and it is well within the abilities of one skilled in the art to adapt these conditions for use with a particular polypeptide hybrid. The T_m for a specific target sequence is the temperature (under defined conditions) at which 50% of the target sequence will hybridize to a perfectly matched probe sequence. Those conditions which influence the T_m include, the size and base pair content of the polynucleotide probe, the ionic strength of the hybridization solution, and the presence of destabilizing agents in the hybridization solution. Numerous equations for calculating T_m are known in the art, see for example (Sambrook et al., *ibid.*; Ausubel et al., *ibid.*; Berger and Kimmel, *ibid.* and Wetmur, *ibid.*) and are specific for DNA, RNA and DNA-RNA hybrids and polynucleotide probe sequences of varying length. Sequence analysis software such as Oligo 4.0 (publicly available shareware) and Primer Premier (PREMIER Biosoft International, Palo Alto, Calif.) as well as sites on the Internet, are available tools for analyzing a given sequence and calculating T_m based on user defined criteria. Such programs can also analyze a given sequence under defined conditions and suggest suitable probe sequences. Typically, hybridization of longer polynucleotide sequences, >50 bp, is done at temperatures of about 20–25° C. below the calculated T_m . For smaller probes, <50 bp, hybridization is typically carried out at the T_m or 5–10° C. below. This allows for the maximum rate of hybridization for DNA-DNA and DNA-RNA hybrids.

The length of the polynucleotide sequence influences the rate and stability of hybrid formation. Smaller probe sequences, <50 bp, come to equilibrium with complementary sequences rapidly, but may form less stable hybrids. Incubation times of anywhere from minutes to hours can be used to achieve hybrid formation. Longer probe sequences come to equilibrium more slowly, but form more stable complexes even at lower temperatures. Incubations are allowed to proceed overnight or longer. Generally, incubations are carried out for a period equal to three times the calculated Cot time. Cot time, the time it takes for the polynucleotide sequences to reassociate, can be calculated for a particular sequence by methods known in the art.

The base pair composition of polynucleotide sequence will effect the thermal stability of the hybrid complex, thereby influencing the choice of hybridization temperature and the ionic strength of the hybridization buffer. A-T pairs

are less stable than G-C pairs in aqueous solutions containing NaCl. Therefore, the higher the G-C content, the more stable the hybrid. Even distribution of G and C residues within the sequence also contribute positively to hybrid stability. Base pair composition can be manipulated to alter the T_m of a given sequence, for example, 5-methyldeoxycytidine can be substituted for deoxycytidine and 5-bromodeoxyuridine can be substituted for thymidine to increase the T_m . 7-deaza-2'-deoxyguanosine can be substituted for guanosine to reduce dependence on T_m .

Ionic concentration of the hybridization buffer also effects the stability of the hybrid. Hybridization buffers generally contain blocking agents such as Denhardt's solution (Sigma Chemical Co., St. Louis, Mo.), denatured salmon sperm DNA, tRNA, milk powders (BLOTTO), heparin or SDS, and a Na⁺ source, such as SSC (1× SSC: 0.15 M NaCl, 15 mM sodium citrate) or SSPE (1× SSPE: 1.8 M NaCl, 10 mM NaH₂PO₄, 1 mM EDTA, pH 7.7). By decreasing the ionic concentration of the buffer, the stability of the hybrid is increased. Typically, hybridization buffers contain from between 10 mM–1 M Na⁺. Premixed hybridization solutions are also available from commercial sources such as Clontech Laboratories (Palo Alto, Calif.) and Promega Corporation (Madison, Wis.) for use according to manufacturer's instruction. Addition of destabilizing or denaturing agents such as formamide, tetraalkylammonium salts, guanidinium cations or thiocyanate cations to the hybridization solution will alter the T_m of a hybrid. Typically, formamide is used at a concentration of up to 50% to allow incubations to be carried out at more convenient and lower temperatures. Formamide also acts to reduce non-specific background when using RNA probes.

As previously noted, the isolated zpep10 polynucleotides of the present invention include DNA and RNA. Methods for isolating DNA and RNA are well known in the art. It is generally preferred to isolate RNA from lymph node, although DNA can also be prepared using RNA from other tissues or isolated as genomic DNA. Total RNA can be prepared using guanidine HCl extraction followed by isolation by centrifugation in a CsCl gradient (Chirgwin et al., *Biochemistry* 18:52–94, 1979). Poly (A)⁺ RNA is prepared from total RNA using the method of Aviv and Leder (*Proc. Natl. Acad. Sci. USA* 69:1408–12, 1972). Complementary DNA (cDNA) is prepared from poly(A)⁺ RNA using known methods. Polynucleotides encoding zpep10 polypeptides are then identified and isolated by, for example, hybridization or PCR.

The polynucleotides of the present invention can also be synthesized using automated equipment. The current method of choice is the phosphoramidite method. If chemically synthesized double stranded DNA is required for an application such as the synthesis of a gene or a gene fragment, then each complementary strand is made separately. The production of short genes (60 to 80 bp) is technically straightforward and can be accomplished by synthesizing the complementary strands and then annealing them. For the production of longer genes (>300 bp), however, special strategies must be invoked, because the coupling efficiency of each cycle during chemical DNA synthesis is seldom 100%. To overcome this problem, synthetic genes (double-stranded) are assembled in modular form from single-stranded fragments that are from 20 to 100 nucleotides in length. Gene synthesis methods are well known in the art. See, for example, Glick and Pasternak, *Molecular Biotechnology, Principles & Applications of Recombinant DNA*, ASM Press, Washington, D.C., 1994; Itakura et al., *Annu. Rev. Biochem.* 53: 323–356, 1984; and Climie et al., *Proc. Natl. Acad. Sci. USA* 87:633–637, 1990.

The zpep10 polynucleotide sequences disclosed herein can be used to isolate polynucleotides encoding other zpep10 proteins. Such other proteins include alternatively spliced cDNAs (including cDNAs encoding secreted zpep10 proteins) and counterpart polynucleotides from other species (orthologs). These orthologous polynucleotides can be used, inter alia, to prepare the respective orthologous proteins. Other species of interest include, but are not limited to, mammalian, avian, amphibian, reptile, fish, insect and other vertebrate and invertebrate species. Of particular interest are zpep10 polynucleotides and proteins from other mammalian species, including human and other primates, porcine, ovine, bovine, canine, feline, and equine polynucleotides and proteins. Orthologs of mouse zpep10, for example, can be cloned using information and compositions provided by the present invention in combination with conventional cloning techniques. For example, a cDNA can be cloned using mRNA obtained from a tissue or cell type that expresses zpep10 as disclosed herein. Suitable sources of mRNA can be identified by probing Northern blots with probes designed from the sequences disclosed herein. A library is then prepared from mRNA of a positive tissue or cell line. A zpep10-encoding cDNA can then be isolated by a variety of methods, such as by probing with a complete or partial human cDNA or with one or more sets of degenerate probes based on the disclosed sequences. A cDNA can also be cloned using the polymerase chain reaction, or PCR (Mullis, U.S. Pat. No. 4,683,202), using primers designed from the representative human zpep10 sequence disclosed herein. Within an additional method, the cDNA library can be used to transform or transfect host cells, and expression of the cDNA of interest can be detected with an antibody to zpep10 polypeptide. Similar techniques can also be applied to the isolation of genomic clones. Electronic databases can also be screened for EST sequences of zpep10 orthologs. Degenerate polynucleotide primer sequences useful for identifying zpep10 orthologs would include:

zpep10 residues 15-20 of SEQ ID NO:2
CARGCNTGYGNTTYTG (SEQ ID NO:4)
zpep10 residues 42-47 of SEQ ID NO:2
CARAARGARTGYGGNGC (SEQ ID NO:5)
zpep10 residues 61-66 of SEQ ID NO:2
ATGAAYAARGRNACNGA (SEQ ID NO:6)
zpep10 residues 64-69 of SEQ ID NO:2
GRNACNGARAARACNCA (SEQ ID NO:7)
zpep10 residues 86-91 of SEQ ID NO:2
ACNTGYAARGGNACNGA (SEQ ID NO:8).

Those skilled in the art will recognize that the sequences disclosed in SEQ ID NO:1 and SEQ ID NO:2 represent a single allele of the human zpep10 gene and polypeptide, and that allelic variation and alternative splicing are expected to occur. In addition, allelic variants can be cloned by probing cDNA or genomic libraries from different individuals according to standard procedures. Allelic variants of the DNA sequence shown in SEQ ID NO:1, including those containing silent mutations and those in which mutations result in amino acid sequence changes, are within the scope of the present invention, as are proteins which are allelic variants of SEQ ID NO:2. cDNAs generated from alternatively spliced mRNAs, which retain the properties of the zpep10 polypeptide are included within the scope of the present invention, as are polypeptides encoded by such cDNAs and mRNAs. Allelic variants and splice variants of these sequences can be cloned by probing cDNA or genomic libraries from different individuals or tissues according to standard procedures known in the art.

The present invention also provides isolated zpep10 polypeptides that are substantially homologous to the polypeptide of SEQ ID NO:2 and its species orthologs. The term "substantially homologous" is used herein to denote polypeptides having 60%, preferably at least 80%, sequence identity to the sequences shown in SEQ ID NO:2 or their orthologs. Such polypeptides will more preferably be at least 90% identical, and most preferably 95% or more identical to SEQ ID NO:2 or its orthologs. Percent sequence identity is determined by conventional methods. See, for example, Altschul et al., *Bull. Math. Bio.* 48: 603-16, 1986 and Henikoff and Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915-9, 1992. Briefly, two amino acid sequences are aligned to optimize the alignment scores using a gap opening penalty of 10, a gap extension penalty of 1, and the "blosum 62" scoring matrix of Henikoff and Henikoff (ibid.) as shown in Table 3 (amino acids are indicated by the standard one-letter codes). The percent identity is then calculated as:

$$\frac{\text{Total number of identical matches}}{[\text{length of the longer sequence plus the number of gaps introduced into the longer sequence in order to align the two sequences}]} \times 100$$

TABLE 3

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	4																			
R	-1	5																		
N	-2	0	6																	
D	-2	-2	1	6																
C	0	-3	-3	-3	9															
Q	-1	1	0	0	-3	5														
E	-1	0	0	2	-4	2	5													
G	0	-2	0	-1	-3	-2	-2	6												
H	-2	0	1	-1	-3	0	0	-2	8											
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4										
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4									
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5								
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5							
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6						
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7					
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4				

TABLE 3-continued

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5			
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11		
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4

Sequence identity of polynucleotide molecules is determined by similar methods using a ratio as disclosed above.

Those skilled in the art appreciate that there are many established algorithms available to align two amino acid sequences. The "FASTA" similarity search algorithm of Pearson and Lipman is a suitable protein alignment method for examining the level of identity shared by an amino acid sequence disclosed herein and the amino acid sequence of a putative variant zpep10. The FASTA algorithm is described by Pearson and Lipman, *Proc. Nat. Acad. Sci. USA* 85:2444, 1988, and by Pearson, *Meth. Enzymol.* 183:63, 1990.

Briefly, FASTA first characterizes sequence similarity by identifying regions shared by the query sequence (e.g., SEQ ID NO:2) and a test sequence that have either the highest density of identities (if the ktup variable is 1) or pairs of identities (if ktup=2), without considering conservative amino acid substitutions, insertions, or deletions. The ten regions with the highest density of identities are then re-scored by comparing the similarity of all paired amino acids using an amino acid substitution matrix, and the ends of the regions are "trimmed" to include only those residues that contribute to the highest score. If there are several regions with scores greater than the "cutoff" value (calculated by a predetermined formula based upon the length of the sequence and the ktup value), then the trimmed initial regions are examined to determine whether the regions can be joined to form an approximate alignment with gaps. Finally, the highest scoring regions of the two amino acid sequences are aligned using a modification of the Needleman-Wunsch-Sellers algorithm (Needleman and Wunsch, *J. Mol. Biol.* 48:444, 1970; Sellers, *SIAM J. Appl. Math.* 26:787, 1974), which allows for amino acid insertions and deletions. Preferred parameters for FASTA analysis are: ktup=1, gap opening penalty=10, gap extension penalty=1, and substitution matrix=BLOSUM62. These parameters can be introduced into a FASTA program by modifying the scoring matrix file ("SMATRIX"), as explained in Appendix 2 of Pearson, *Meth. Enzymol.* 183:63, 1990.

FASTA can also be used to determine the sequence identity of nucleic acid molecules using a ratio as disclosed above. For nucleotide sequence comparisons, the ktup value can range between one to six, preferably from three to six, most preferably three, with other parameters set as default.

The BLOSUM62 table is an amino acid substitution matrix derived from about 2,000 local multiple alignments of protein sequence segments, representing highly conserved regions of more than 500 groups of related proteins (Henikoff and Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915, 1992). Accordingly, the BLOSUM62 substitution frequencies can be used to define conservative amino acid substitutions that may be introduced into the amino acid sequences of the present invention. Although it is possible to design amino acid substitutions based solely upon chemical properties (as discussed above), the language "conservative amino acid substitution" preferably refers to a substitution represented by a BLOSUM62 value of greater than -1. For example, an amino acid substitution is conservative if the substitution is characterized by a BLOSUM62 value of 0, 1,

2, or 3. According to this system, preferred conservative amino acid substitutions are characterized by a BLOSUM62 value of at least 1 (e.g., 1, 2 or 3), while more preferred conservative amino acid substitutions are characterized by a BLOSUM62 value of at least 2 (e.g., 2 or 3).

Substantially homologous proteins and polypeptides are characterized as having one or more amino acid substitutions, deletions or additions. These changes are preferably of a minor nature, that is conservative amino acid substitutions (see Table 4) and other substitutions that do not significantly affect the folding or activity of the protein or polypeptide; small deletions, typically of one to about 30 amino acids; and small amino- or carboxyl-terminal extensions, such as an amino-terminal methionine residue, a small linker peptide of up to about 20-25 residues or an affinity tag. Polypeptides comprising affinity tags can further comprise a proteolytic cleavage site between the zpep10 polypeptide and the affinity tag. Preferred such sites include thrombin cleavage sites and factor Xa cleavage sites.

TABLE 4

Conservative amino acid substitutions	
Basic:	arginine lysine histidine
Acidic:	glutamic acid aspartic acid
Polar:	glutamine asparagine
Hydrophobic:	leucine isoleucine valine
Aromatic:	phenylalanine tryptophan tyrosine
Small:	glycine alanine serine threonine methionine

The proteins of the present invention can also comprise non-naturally occurring amino acid residues. Non-naturally occurring amino acids include, without limitation, trans-3-methylproline, 2,4-methanoproline, cis-4-hydroxyproline, trans-4-hydroxyproline, N-methyl-glycine, allo-threonine, methylthreonine, hydroxyethyl-cysteine, hydroxyethylhomocysteine, nitroglutamine, homoglutamine, pipecolic acid, thiazolidine carboxylic acid, dehydropyrolidine, 3- and 4-methylproline, 3,3-dimethylproline, tert-leucine, norvaline, 2-azaphenyl-alanine, 3-azaphenylalanine, 4-azaphenylalanine, and 4-fluorophenyl-alanine. Several methods are known in the art for incorporating non-naturally occurring amino acid residues into proteins. For example, an in vitro system can be employed wherein nonsense mutations are suppressed using chemically aminoacylated suppressor tRNAs. Methods for synthesizing amino acids and aminoacylating tRNA are known in the art. Transcription and translation of plas-

mids containing nonsense mutations is carried out in a cell-free system comprising an *E. coli* S30 extract and commercially available enzymes and other reagents. Proteins are purified by chromatography. See, for example, Robertson et al., *J. Am. Chem. Soc.* 113:2722, 1991; Ellman et al., *Methods Enzymol.* 202:301, 1991; Chung et al., *Science* 259:806-9, 1993; and Chung et al., *Proc. Natl. Acad. Sci. USA* 90:10145-9, 1993). In a second method, translation is carried out in *Xenopus* oocytes by microinjection of mutated mRNA and chemically aminoacylated suppressor tRNAs (Turcatti et al., *J. Biol. Chem.* 271:19991-8, 1996). Within a third method, *E. coli* cells are cultured in the absence of a natural amino acid that is to be replaced (e.g., phenylalanine) and in the presence of the desired non-naturally occurring amino acid(s) (e.g., 2-azaphenylalanine, 3-azaphenylalanine, 4-azaphenylalanine, or 4-fluorophenylalanine). The non-naturally occurring amino acid is incorporated into the protein in place of its natural counterpart. See, Koide et al., *Biochem.* 33:7470-6, 1994. Naturally occurring amino acid residues can be converted to non-naturally occurring species by in vitro chemical modification. Chemical modification can be combined with site-directed mutagenesis to further expand the range of substitutions (Wynn and Richards, *Protein Sci.* 2:39-403, 1993).

A limited number of non-conservative amino acids, amino acids that are not encoded by the genetic code, non-naturally occurring amino acids, and unnatural amino acids may be substituted for zpep10 amino acid residues.

Essential amino acids in the zpep10 polypeptides of the present invention can be identified according to procedures known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham and Wells, *Science* 244: 1081-5, 1989). In the latter technique, single alanine mutations are introduced at every residue in the molecule, and the resultant mutant molecules are tested for biological activity (e.g., adhesion-modulation, differentiation-modulation or the like) to identify amino acid residues that are critical to the activity of the molecule. See also, Hilton et al., *J. Biol. Chem.* 271:4699-708, 1996. Sites of ligand-receptor or other biological interaction can also be determined by physical analysis of structure, as determined by such techniques as nuclear magnetic resonance, crystallography, electron diffraction or photoaffinity labeling, in conjunction with mutation of putative contact site amino acids. See, for example, de Vos et al., *Science* 255:306-12, 1992; Smith et al., *J. Mol. Biol.* 224:899-904, 1992; Wlodaver et al., *FEBS Lett.* 309:59-64, 1992. The identities of essential amino acids can also be inferred from analysis of homologies with related proteins. Amino acid residues that might be considered essential in the zpep10 polypeptide are cysteine residues at amino acid residues 17, 20, 35, 45, 84, 87, 94 and 100 of SEQ ID NO:2; the potential arg-lys-arg tri-basic amino acid cleavage site at amino acid residues 97-99 of SEQ ID NO:2; the potential N-linked glycosylation sites at amino acid residues 83 and 86 of SEQ ID NO:2 and the potential O-glycosylation sites at amino acid residues 28, 36, 48, 52, 60, 65, 68, 78, 79, 80, 85, 86, 90, 93 and 104 of SEQ ID NO:2. A hydrophobicity profile of SEQ ID NO:2 is shown in the attached FIGURE. Those skilled in the art will recognize that this hydrophobicity will be taken into account when designing alterations in the amino acid sequence of a zpep10 polypeptide, so as not to disrupt the overall profile.

Multiple amino acid substitutions can be made and tested using known methods of mutagenesis and screening, such as those disclosed by Reidhaar-Olson and Sauer (*Science*

241:53-57, 1988) or Bowie and Sauer (*Proc. Natl. Acad. Sci. USA* 86:2152-2156, 1989). Briefly, these authors disclose methods for simultaneously randomizing two or more positions in a polypeptide, selecting for functional polypeptide, and then sequencing the mutagenized polypeptides to determine the spectrum of allowable substitutions at each position. Other methods that can be used include phage display (e.g., Lowman et al., *Biochem.* 30:10832-10837, 1991; Ladner et al., U.S. Pat. No. 5,223,409; Huse, WIPO Publication WO 92/06204) and region-directed mutagenesis (Derbyshire et al., *Gene* 46:145, 1986; Ner et al., *DNA* 7:127, 1988).

Variants of the disclosed zpep10 DNA and polypeptide sequences can be generated through DNA shuffling as disclosed by Stemmer, *Nature* 370:389-91, 1994, Stemmer, *Proc. Natl. Acad. Sci. USA* 91:10747-51, 1994 and WIPO Publication WO 97/20078. Briefly, variant DNAs are generated by in vitro homologous recombination by random fragmentation of a parent DNA followed by reassembly using PCR, resulting in randomly introduced point mutations. This technique can be modified by using a family of parent DNAs, such as allelic variants or DNAs from different species, to introduce additional variability into the process. Selection or screening for the desired activity, followed by additional iterations of mutagenesis and assay provides for rapid "evolution" of sequences by selecting for desirable mutations while simultaneously selecting against detrimental changes.

Mutagenesis methods as disclosed above can be combined with high-throughput, automated screening methods to detect activity of cloned, mutagenized polypeptides in host cells. Mutagenized DNA molecules that encode active polypeptides (e.g., ligand binding receptors) can be recovered from the host cells and rapidly sequenced using modern equipment. These methods allow the rapid determination of the importance of individual amino acid residues in a polypeptide of interest, and can be applied to polypeptides of unknown structure.

Using the methods discussed above, one of ordinary skill in the art can identify and/or prepare a variety of polypeptides that are substantially homologous to, for example, residues 21 to 111, 21 to 142 or 1 to 142 of SEQ ID NO:2 or allelic variants thereof and retain the properties of wild-type protein. Such polypeptides may include additional amino acids, such as affinity tags and the like. Such polypeptides may also include additional polypeptide segments as generally disclosed herein.

The invention also provides soluble polypeptides. It is preferred that these soluble polypeptides be extracellular polypeptides and be in a form substantially free of transmembrane and intracellular polypeptide segments. To direct the export of the soluble polypeptides from the host cell, the DNA encoding the soluble polypeptide is linked to a second DNA segment encoding a secretory peptide, such as a t-PA secretory peptide or the native zpep10 secretory signal sequence (amino acid residues 1-20 of SEQ ID NO:2). To facilitate purification of the secreted polypeptide, an N- or C-terminal extension, such as an affinity tag or another polypeptide or protein for which an antibody or other specific binding agent is available, can be fused to the soluble polypeptide.

The present invention also provides zpep10fusion proteins. For example, fusion proteins of the present invention encompass

- (1) a polypeptide selected from the following: a) a polypeptide comprising a sequence of amino acid residues from amino acid residue 21 to amino acid residue

111 of SEQ ID NO:2; and b) a polypeptide comprising a sequence of amino acid residues from amino acid residue 1 to amino acid residue 20 of SEQ ID NO:2; and

- (2) another polypeptide. The other polypeptide may be a signal peptide to facilitate secretion of the fusion protein, a transmembrane and/or cytoplasmic domain, or another soluble polypeptide or the like. For example, the extracellular portion of a zpep10 polypeptide can be prepared as a fusion to a dimerizing protein as disclosed in U.S. Pat. Nos. 5,155,027 and 5,567,584. Preferred dimerizing proteins in this regard include immunoglobulin constant region domains. Immunoglobulin-zpep10 polypeptide fusions can be expressed in genetically engineered cells to produce a variety of multimeric zpep10 analogs. Auxiliary domains can be fused to zpep10 polypeptides to target them to specific cells, tissues, or macromolecules. For example, a soluble zpep10 polypeptide or protein could be targeted to a predetermined cell type by fusing a zpep10 polypeptide to a ligand that specifically binds to a receptor on the surface of the target cell. In this way, polypeptides and proteins can be targeted for therapeutic or diagnostic purposes. A zpep10 polypeptide can be fused to two or more moieties, such as an affinity tag for purification and a targeting domain. Polypeptide fusions can also comprise one or more cleavage sites, particularly between domains. See, Tuan et al., *Connective Tissue Research* 34:1-9, 1996.

The soluble zpep10 polypeptide is useful in studying the distribution of zpep10 receptors on tissues or specific cell lineages, and to provide insight into receptor/ligand biology. Using labeled soluble zpep10, cells expressing the ligand are identified by fluorescence immunocytometry or immunohistochemistry. The effects of zpep10 on steroidogenesis or Leydig or Sertoli cell expression can be examined by probing tissue slices with soluble zpep10 fusions, see for example, Dachlin et al., *Scand. J. Urol. Nephrol.* 19:7-12, 1985; Gavino et al., *Arch. Biochem. Biophys.* 233:741-7, 1984 and von Schnakenburg et al., *Acta Endocrinol.* 94:397-403, 1980). luteinizing hormone (LH) and follicle stimulating hormone (FSH) responses could also be examined in soluble zpep10-treated tissue slices.

The polypeptides of the present invention, including full-length proteins, fragments thereof and fusion proteins, can be produced in genetically engineered host cells according to conventional techniques. Suitable host cells are those cell types that can be transformed or transfected with exogenous DNA and grown in culture, and include bacteria, fungal cells, and cultured higher eukaryotic cells. Eukaryotic cells, particularly cultured cells of multicellular organisms, are preferred. Techniques for manipulating cloned DNA molecules and introducing exogenous DNA into a variety of host cells are disclosed by Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989, and Ausubel et al. (eds.), *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc., NY, 1987.

In general, a DNA sequence encoding a zpep10 polypeptide of the present invention is operably linked to other genetic elements required for its expression, generally including a transcription promoter and terminator within an expression vector. The vector will also commonly contain one or more selectable markers and one or more origins of replication, although those skilled in the art will recognize that within certain systems selectable markers may be provided on separate vectors, and replication of the exogenous

DNA may be provided by integration into the host cell genome. Selection of promoters, terminators, selectable markers, vectors and other elements is a matter of routine design within the level of ordinary skill in the art. Many such elements are described in the literature and are available through commercial suppliers.

To direct a zpep10 polypeptide into the secretory pathway of a host cell, a secretory signal sequence (also known as a signal sequence, leader sequence, prepro sequence or pre sequence) is provided in the expression vector. The secretory signal sequence may be that of the zpep10 polypeptide, or may be derived from another secreted protein (e.g., t-PA) or synthesized de novo. The secretory signal sequence is joined to the zpep10 DNA sequence in the correct reading frame and positioned to direct newly synthesized polypeptide into secretory pathways to host cell. Secretory signal sequences are commonly positioned 5' to the DNA sequence encoding the polypeptide of interest, although certain secretory signal sequences may be positioned elsewhere in the DNA sequence of interest (see, e.g., Welch et al., U.S. Pat. No. 5,037,743; Holland et al., U.S. Pat. No. 5,143,830).

Alternatively, the secretory signal sequence contained in the polypeptides of the present invention is used to direct other polypeptides into the secretory pathway. The present invention provides for such fusion polypeptides. A signal fusion polypeptide can be made wherein a secretory signal sequence derived from amino acid residues 1-20 of SEQ ID NO:2 is operably linked to another polypeptide using methods known in the art and disclosed herein. The secretory signal sequence contained in the fusion polypeptides of the present invention is preferably fused amino-terminally to an additional peptide to direct the additional peptide into the secretory pathway. Such constructs have numerous applications known in the art. For example, these novel secretory signal sequence fusion constructs can direct the secretion of an active component of a normally non-secreted protein, such as a receptor. Such fusions may be used in vivo or in vitro to direct peptides through the secretory pathway.

Cultured mammalian cells are suitable hosts within the present invention. Methods for introducing exogenous DNA into mammalian host cells include calcium phosphate-mediated transfection (Wigler et al., *Cell* 14:725, 1978; Corsaro and Pearson, *Somatic Cell Genetics* 7:603, 1981; Graham and Van der Eb, *Virology* 52:456, 1973), electroporation (Neumann et al., *EMBO J.* 1:841-845, 1982), DEAE-dextran mediated transfection (Ausubel et al., eds., *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc., NY, 1987), liposome-mediated transfection (Hawley-Nelson et al., *Focus* 15:73, 1993; Ciccarone et al., *Focus* 15:80, 1993), and viral vectors (Miller and Rosman, *BioTechniques* 7:980-90, 1989; Wang and Finer, *Nature Med.* 2:714-16, 1996). The production of recombinant polypeptides in cultured mammalian cells is disclosed, for example, by Levinson et al., U.S. Pat. No. 4,713,339; Hagen et al., U.S. Pat. No. 4,784,950; Palmiter et al., U.S. Pat. No. 4,579,821; and Ringold, U.S. Pat. No. 4,656,134. Suitable cultured mammalian cells include the COS-1 (ATCC No. CRL 1650), COS-7 (ATCC No. CRL 1651), BHK 570 (ATCC No. CRL 10314), 293 (ATCC No. CRL 1573; Graham et al., *J. Gen. Virol.* 36:59-72, 1977) and Chinese hamster ovary (e.g. CHO-K1; ATCC No. CCL 61) cell lines. Additional suitable cell lines are known in the art and available from public depositories such as the American Type Culture Collection, Rockville, Md. In general, strong transcription promoters are preferred, such as promoters from SV-40 or cytomegalovirus. See, e.g., U.S. Pat. No. 4,956,288. Other suitable promoters include those from metallothionein genes (U.S.

Pat. Nos. 4,579,821 and 4,601,978) and the adenovirus major late promoter.

Drug selection is generally used to select for cultured mammalian cells into which foreign DNA has been inserted. Such cells are commonly referred to as "transfectants". Cells that have been cultured in the presence of the selective agent and are able to pass the gene of interest to their progeny are referred to as "stable transfectants." A preferred selectable marker is a gene encoding resistance to the antibiotic neomycin. Selection is carried out in the presence of a neomycin-type drug, such as G-418 or the like. Selection systems may also be used to increase the expression level of the gene of interest, a process referred to as "amplification." Amplification is carried out by culturing transfectants in the presence of a low level of the selective agent and then increasing the amount of selective agent to select for cells that produce high levels of the products of the introduced genes. A preferred amplifiable selectable marker is dihydrofolate reductase, which confers resistance to methotrexate. Other drug resistance genes (e.g., hygromycin resistance, multi-drug resistance, puromycin acetyltransferase) can also be used. Alternative markers that introduce an altered phenotype, such as green fluorescent protein, or cell surface proteins such as CD4, CD8, Class I MHC, placental alkaline phosphatase may be used to sort transfected cells from untransfected cells by such means as FACS sorting or magnetic bead separation technology.

Other higher eukaryotic cells can also be used as hosts, including plant cells, insect cells and avian cells. The use of *Agrobacterium rhizogenes* as a vector for expressing genes in plant cells has been reviewed by Sinkar et al., *J. Biosci.* (Bangalore) 11:47-58, 1987. Transformation of insect cells and production of foreign polypeptides therein is disclosed by Guarino et al., U.S. Pat. No. 5,162,222 and WIPO publication WO 94/06463. Insect cells can be infected with recombinant baculovirus; commonly derived from *Autographa californica* nuclear polyhedrosis virus (AcNPV). DNA encoding the zpep10 polypeptide is inserted into the baculoviral genome in place of the AcNPV polyhedrin gene coding sequence by one of two methods. The first is the traditional method of homologous DNA recombination between wild-type AcNPV and a transfer vector containing the zpep10 flanked by AcNPV sequences. Suitable insect cells, e.g. Sf9 cells, are infected with wild-type AcNPV and transfected with a transfer vector comprising a zpep10 polynucleotide operably linked to an AcNPV polyhedrin gene promoter, terminator, and flanking sequences. See, King and Possee, *The Baculovirus Expression System: A Laboratory Guide*, London, Chapman & Hall; O'Reilly et al., *Baculovirus Expression Vectors: A Laboratory Manual*, New York, Oxford University Press., 1994; and, Richardson, Ed., *Baculovirus Expression Protocols. Methods in Molecular Biology*, Totowa, N.J., Humana Press, 1995. Natural recombination within an insect cell will result in a recombinant baculovirus which contains zpep10 driven by the polyhedrin promoter. Recombinant viral stocks are made by methods commonly used in the art.

The second method of making recombinant baculovirus utilizes a transposon-based system described by Luckow et al. (*J. Virol.* 67:4566-79, 1993). This system is sold in the Bac-to-Bac kit (Life Technologies, Rockville, Md.). This system utilizes a transfer vector, pFastBac1™ (Life Technologies) containing a Tn7 transposon to move the DNA encoding the zpep10 polypeptide into a baculovirus genome maintained in *E. coli* as a large plasmid called a "bacmid." The pFastBac1™ transfer vector utilizes the AcNPV polyhedrin promoter to drive the expression of the

gene of interest, in this case zpep10. However, pFastBac1™ can be modified to a considerable degree. The polyhedrin promoter can be removed and substituted with the baculovirus basic protein promoter (also known as Pcor, p6.9 or MP promoter) which is expressed earlier in the baculovirus infection, and has been shown to be advantageous for expressing secreted proteins. See, Hill-Perkins and Possee, *J. Gen. Virol.* 71:971-6, 1990; Bonning et al., *J. Gen. Virol.* 75:1551-6, 1994; and, Chazenbalk and Rapoport, *J. Biol. Chem.* 270:1543-9, 1995. In such transfer vector constructs, a short or long version of the basic protein promoter can be used. Moreover, transfer vectors can be constructed which replace the native zpep10 secretory signal sequences with secretory signal sequences derived from insect proteins. For example, a secretory signal sequence from Ecdysteroid Glucosyltransferase (EGT), honey bee Melittin (Invitrogen, Carlsbad, Calif.), or baculovirus gp67 (PharMingen, San Diego, Calif.) can be used in constructs to replace the native secretory signal sequence. In addition, transfer vectors can include an in-frame fusion with DNA encoding an epitope tag at the C- or N-terminus of the expressed zpep10 polypeptide, for example, a Glu-Glu epitope tag (Grussenmeyer et al., *ibid.*) or FLAG tag (Kodak). Using a technique known in the art, a transfer vector containing zpep10 is transformed into *E. coli*, and screened for bacmids which contain an interrupted lacZ gene indicative of recombinant baculovirus. The bacmid DNA containing the recombinant baculovirus genome is isolated, using common techniques, and used to transfect *Spodoptera frugiperda* cells, e.g. Sf9 cells. Recombinant virus that expresses zpep10 is subsequently produced. Recombinant viral stocks are made by methods commonly used in the art.

The recombinant virus is used to infect host cells, typically a cell line derived from the fall armyworm, *Spodoptera frugiperda*. See, in general, Glick and Pasternak, *Molecular Biotechnology: Principles and Applications of Recombinant DNA*, ASM Press, Washington, D.C., 1994. Another suitable cell line is the High FiveO™ cell line (Invitrogen) derived from *Trichoplusia ni* (U.S. Pat. No. 5,300,435). Commercially available serum-free media are used to grow and maintain the cells. Suitable media are Sf900 II™ (Life Technologies) or ESF 921™ (Expression Systems) for the Sf9 cells; and Ex-cello0405™ (JRH Biosciences, Lenexa, Kans.) or Express FiveO™ (Life Technologies) for the T. ni cells. The cells are grown up from an inoculation density of approximately $2-5 \times 10^5$ cells to a density of $1-2 \times 10^6$ cells at which time a recombinant viral stock is added at a multiplicity of infection (MOI) of 0.1 to 10, more typically near 3. The recombinant virus-infected cells typically produce the recombinant zpep10 polypeptide at 12-72 hours post-infection and secrete it with varying efficiency into the medium. The culture is usually harvested 48 hours post-infection. Centrifugation is used to separate the cells from the medium (supernatant). The supernatant containing the zpep10 polypeptide is filtered through micropore filters, usually 0.45 μ m pore size. Procedures used are generally described in available laboratory manuals (King and Possee, *ibid.*; O'Reilly et al., *ibid.*; Richardson, C. D., *ibid.*). Subsequent purification of the zpep10 polypeptide from the supernatant can be achieved using methods described herein.

Fungal cells, including yeast cells, can also be used within the present invention. Yeast species of particular interest in this regard include *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Pichia methanolica*. Methods for transforming *S. cerevisiae* cells with exogenous DNA and producing recombinant polypeptides therefrom are disclosed by, for

example, Kawasaki, U.S. Pat. No. 4,599,311; Kawasaki et al., U.S. Pat. No. 4,931,373; Brake, U.S. Pat. No. 4,870,008; Welch et al., U.S. Pat. No. 5,037,743; and Murray et al., U.S. Pat. No. 4,845,075. Transformed cells are selected by phenotype determined by the selectable marker, commonly drug resistance or the ability to grow in the absence of a particular nutrient (e.g., leucine). A preferred vector system for use in *Saccharomyces cerevisiae* is the POT1 vector system disclosed by Kawasaki et al. (U.S. Pat. No. 4,931,373), which allows transformed cells to be selected by growth in glucose-containing media. Suitable promoters and terminators for use in yeast include those from glycolytic enzyme genes (see, e.g., Kawasaki, U.S. Pat. No. 4,599,311; Kingsman et al., U.S. Pat. No. 4,615,974; and Bitter, U.S. Pat. No. 4,977,092) and alcohol dehydrogenase genes. See also U.S. Pat. Nos. 4,990,446; 5,063,154; 5,139,936 and 4,661,454. Transformation systems for other yeasts, including *Hansenula polymorpha*, *Schizosaccharomyces pombe*, *Kluyveromyces lactis*, *Kluyveromyces fragilis*, *Ustilago maydis*, *Pichia pastoris*, *Pichia methanolica*, *Pichia guillermoidii* and *Candida maltosa* are known in the art. See, for example, Gleeson et al., *J. Gen. Microbiol.* 132:3459-65, 1986 and Cregg, U.S. Pat. No. 4,882,279. *Aspergillus* cells may be utilized according to the methods of McKnight et al., U.S. Pat. No. 4,935,349. Methods for transforming *Acremonium chrysogenum* are disclosed by Sumino et al., U.S. Pat. No. 5,162,228. Methods for transforming *Neurospora* are disclosed by Lambowitz, U.S. Pat. No. 4,486,533.

The use of *Pichia methanolica* as host for the production of recombinant proteins is disclosed in WIPO Publications WO 97/17450, WO 97/17451, WO 98/02536, and WO 98/02565. DNA molecules for use in transforming *P. methanolica* will commonly be prepared as double-stranded, circular plasmids, which are preferably linearized prior to transformation. For polypeptide production in *P. methanolica*, it is preferred that the promoter and terminator in the plasmid be that of a *P. methanolica* gene, such as a *P. methanolica* alcohol utilization gene (AUG1 or AUG2). Other useful promoters include those of the dihydroxyacetone synthase (DHAS), formate dehydrogenase (FMD), and catalase (CAT) genes. To facilitate integration of the DNA into the host chromosome, it is preferred to have the entire expression segment of the plasmid flanked at both ends by host DNA sequences. A preferred selectable marker for use in *Pichia methanolica* is a *P. methanolica* ADE2 gene, which encodes phosphoribosyl-5-aminoimidazole carboxylase (AIRC; EC 4.1.1.21), which allows ade2 host cells to grow in the absence of adenine. For large-scale, industrial processes where it is desirable to minimize the use of methanol, it is preferred to use host cells in which both methanol utilization genes (AUG1 and AUG2) are deleted. For production of secreted proteins, host cells deficient in vacuolar protease genes (PEP4 and PRB1) are preferred. Electroporation is used to facilitate the introduction of a plasmid containing DNA encoding a polypeptide of interest into *P. methanolica* cells. It is preferred to transform *P. methanolica* cells by electroporation using an exponentially decaying, pulsed electric field having a field strength of from 2.5 to 4.5 kV/cm, preferably about 3.75 kV/cm, and a time constant (τ) of from 1 to 40 milliseconds, most preferably about 20 milliseconds.

Prokaryotic host cells, including strains of the bacteria *Escherichia*, *Bacillus* and other genera are also useful host cells within the present invention. Techniques for transforming these hosts and expressing foreign DNA sequences cloned therein are well known in the art (see, e.g., Sambrook et al., *ibid.*). When expressing a zpep10 polypeptide in

bacteria such as *E. coli*, the polypeptide may be retained in the cytoplasm, typically as insoluble granules, or may be directed to the periplasmic space by a bacterial secretion sequence. In the former case, the cells are lysed, and the granules are recovered and denatured using, for example, guanidine isothiocyanate or urea. The denatured polypeptide can then be refolded and dimerized by diluting the denaturant, such as by dialysis against a solution of urea and a combination of reduced and oxidized glutathione, followed by dialysis against a buffered saline solution. In the latter case, the polypeptide can be recovered from the periplasmic space in a soluble and functional form by disrupting the cells (by, for example, sonication or osmotic shock) to release the contents of the periplasmic space and recovering the protein, thereby obviating the need for denaturation and refolding.

Transformed or transfected host cells are cultured according to conventional procedures in a culture medium containing nutrients and other components required for the growth of the chosen host cells. A variety of suitable media, including defined media and complex media, are known in the art and generally include a carbon source, a nitrogen source, essential amino acids, vitamins and minerals. Media may also contain such components as growth factors or serum, as required. The growth medium will generally select for cells containing the exogenously added DNA by, for example, drug selection or deficiency in an essential nutrient which is complemented by the selectable marker carried on the expression vector or co-transfected into the host cell. *P. methanolica* cells are cultured in a medium comprising adequate sources of carbon, nitrogen and trace nutrients at a temperature of about 25° C. to 35° C. Liquid cultures are provided with sufficient aeration by conventional means, such as shaking of small flasks or sparging of fermentors. A preferred culture medium for *P. methanolica* is YEPD (2% D-glucose, 2% Bacto™ Peptone (Difco Laboratories, Detroit, Mich.), 1% Bacto™ yeast extract (Difco Laboratories), 0.004% adenine and 0.006% L-leucine).

Zpep10 polypeptides or fragments thereof may also be prepared through chemical synthesis. Zpep10 polypeptides may be monomers or multimers; glycosylated or non-glycosylated; pegylated or non-pegylated; and may or may not include an initial methionine amino acid residue.

Expressed recombinant zpep10 polypeptides (or chimeric zpep10 polypeptides) can be purified using fractionation and/or conventional purification methods and media. Ammonium sulfate precipitation and acid or chaotrope extraction may be used for fractionation of samples. Exemplary purification steps may include hydroxyapatite, size exclusion, FPLC and reverse-phase high performance liquid chromatography. Suitable anion exchange media include derivatized dextrans, agarose, cellulose, polyacrylamide, specialty silicas, and the like. DEAE Fast-Flow Sepharose (Pharmacia, Piscataway, N.J.), PEI, DEAE, QAE and Q derivatives are preferred. Exemplary chromatographic media include those media derivatized with phenyl, butyl, or octyl groups, such as Phenyl-Sepharose FF (Pharmacia), Toyopearl butyl 650 (Toso Haas, Montgomeryville, Pa.), Octyl-Sepharose (Pharmacia) and the like; or polyacrylic resins, such as Amberchrom CG 71 (Toso Haas) and the like. Suitable solid supports include glass beads, silica-based resins, cellulosic resins, agarose beads, cross-linked agarose beads, polystyrene beads, cross-linked polyacrylamide resins and the like that are insoluble under the conditions in which they are to be used. These supports may be modified with reactive groups that allow attachment of proteins by amino groups, carboxyl groups, sulfhydryl groups, hydroxyl

groups and/or carbohydrate moieties. Examples of coupling chemistries include cyanogen bromide activation, N-hydroxysuccinimide activation, epoxide activation, sulfhydryl activation, hydrazide activation, and carboxyl and amino derivatives for carbodiimide coupling chemistries. These and other solid media are well known and widely used in the art, and are available from commercial suppliers. Methods for binding receptor polypeptides to support media are well known in the art. Selection of a particular method is a matter of routine design and is determined in part by the properties of the chosen support. See, for example, *Affinity Chromatography: Principles & Methods*, Pharmacia LKB Biotechnology, Uppsala, Sweden, 1988.

The zpep10 polypeptides of the present invention can be isolated by exploitation of their structural features. Within one embodiment of the invention are included a fusion of the polypeptide of interest and an affinity tag (e.g., polyhistidine, Glu—Glu, FLAG, maltose-binding protein, an immunoglobulin domain) that may be constructed to facilitate purification.

Protein refolding (and optionally reoxidation) procedures may be advantageously used. It is preferred to purify the protein to >80% purity, more preferably to >90% purity, even more preferably >95%, and particularly preferred is a pharmaceutically pure state, that is greater than 99.9% pure with respect to contaminating macromolecules, particularly other proteins and nucleic acids, and free of infectious and pyrogenic agents. Preferably, a purified protein is substantially free of other proteins, particularly other proteins of animal origin.

Proteins/polypeptides which bind zpep10 (such as a zpep10-binding receptor or other membrane glycoprotein) can also be used for purification of zpep10. The zpep10-binding protein/polypeptide is immobilized on a solid support, such as beads of agarose, cross-linked agarose, glass, cellulosic resins, silica-based resins, polystyrene, cross-linked polyacrylamide, or like materials that are stable under the conditions of use. Methods for linking polypeptides to solid supports are known in the art, and include amine chemistry, cyanogen bromide activation, N-hydroxysuccinimide activation, epoxide activation, sulfhydryl activation, and hydrazide activation. The resulting medium will generally be configured in the form of a column, and fluids containing zpep10 polypeptide are passed through the column one or more times to allow zpep10 polypeptide to bind to the ligand-binding or receptor polypeptide. The bound zpep10 polypeptide is then eluted using changes in salt concentration, chaotropic agents (guanidine HCl), or pH to disrupt ligand-receptor binding.

In vitro and in vivo response to soluble zpep10 can also be measured using cultured cells or by administering molecules of the claimed invention to the appropriate animal model. For instance, soluble zpep10 transfected expression host cells may be embedded in an alginate environment and injected (implanted) into recipient animals. Alginate-poly-L-lysine micro-encapsulation, permselective membrane encapsulation and diffusion chambers have been described as a means to entrap transfected mammalian cells or primary mammalian cells. These types of non-immunogenic "encapsulations" or microenvironments permit the transfer of nutrients into the microenvironment, and also permit the diffusion of proteins and other macromolecules secreted or released by the captured cells across the environmental barrier to the recipient animal. Most importantly, the capsules or microenvironments mask and shield the foreign, embedded cells from the recipient animal's immune response. Such microenvironments can extend the life of the

injected cells from a few hours or days (naked cells) to several weeks (embedded cells).

Alginate threads provide a simple and quick means for generating embedded cells. The materials needed to generate the alginate threads are readily available and relatively inexpensive. Once made, the alginate threads are relatively strong and durable, both in vitro and, based on data obtained using the threads, in vivo. The alginate threads are easily manipulable and the methodology is scalable for preparation of numerous threads. In an exemplary procedure, 3% alginate is prepared in sterile H₂O, and sterile filtered. Just prior to preparation of alginate threads, the alginate solution is again filtered. An approximately 50% cell suspension (containing about 5×10^5 to about 5×10^7 cells/ml) is mixed with the 3% alginate solution. One ml of the alginate/cell suspension is extruded into a 100 mM sterile filtered CaCl₂ solution over a time period of ~15 min, forming a "thread". The extruded thread is then transferred into a solution of 50 mM CaCl₂, and then into a solution of 25 mM CaCl₂. The thread is then rinsed with deionized water before coating the thread by incubating in a 0.01% solution of poly-L-lysine. Finally, the thread is rinsed with Lactated Ringer's Solution and drawn from solution into a syringe barrel (without needle attached). A large bore needle is then attached to the syringe, and the thread is intraperitoneally injected into a recipient in a minimal volume of the Lactated Ringer's Solution.

An alternative in vivo approach for assaying soluble proteins of the present invention involves viral delivery systems. Exemplary viruses for this purpose include adenovirus, herpesvirus, vaccinia virus and adeno-associated virus (AAV). Adenovirus, a double-stranded DNA virus, is currently the best studied gene transfer vector for delivery of heterologous nucleic acid (for a review, see Becker et al., *Meth. Cell Biol.* 43:161-89, 1994; and Douglas and Curiel, *Science & Medicine* 4:44-53, 1997). The adenovirus system offers several advantages: adenovirus can (i) accommodate relatively large DNA inserts; (ii) be grown to high-titer; (iii) infect a broad range of mammalian cell types; and (iv) be used with a large number of available vectors containing different promoters. Also, because adenoviruses are stable in the bloodstream, they can be administered by intravenous injection. Some disadvantages (especially for gene therapy) associated with adenovirus gene delivery include: (i) very low efficiency integration into the host genome; (ii) existence in primarily episomal form; and (iii) the host immune response to the administered virus, precluding readministration of the adenoviral vector.

By deleting portions of the adenovirus genome, larger inserts (up to 7 kb) of heterologous DNA can be accommodated. These inserts can be incorporated into the viral DNA by direct ligation or by homologous recombination with a co-transfected plasmid. In an exemplary system, the essential E1 gene has been deleted from the viral vector, and the virus will not replicate unless the E1 gene is provided by the host cell (the human 293 cell line is exemplary). When intravenously administered to intact animals, adenovirus primarily targets the liver. If the adenoviral delivery system has an E1 gene deletion, the virus cannot replicate in the host cells. However, the host's tissue (e.g., liver) will express and process (and, if a secretory signal sequence is present, secrete) the heterologous protein. Secreted proteins will enter the circulation in the highly vascularized liver, and effects on the infected animal can be determined.

The adenovirus system can also be used for protein production in vitro. By culturing adenovirus-infected non-293 cells under conditions where the cells are not rapidly

dividing, the cells can produce proteins for extended periods of time. For instance, BHK cells are grown to confluence in cell factories, then exposed to the adenoviral vector encoding the secreted protein of interest. The cells are then grown under serum-free conditions, which allows infected cells to survive for several weeks without significant cell division. Alternatively, adenovirus vector infected 293S cells can be grown in suspension culture at relatively high cell density to produce significant amounts of protein (see Garnier et al., *Cytotechnol.* 15:145-55, 1994). With either protocol, an expressed, secreted heterologous protein can be repeatedly isolated from the cell culture supernatant. Within the infected 293S cell production protocol, non-secreted proteins may also be effectively obtained.

An assay system that uses a ligand-binding receptor (or an antibody, one member of a complement/anti-complement pair) or a binding fragment thereof, and a commercially available biosensor instrument (BIAcore™, Pharmacia Biosensor, Piscataway, N.J.) may be advantageously employed. Such receptor, antibody, member of a complement/anti-complement pair or fragment is immobilized onto the surface of a receptor chip. Use of this instrument is disclosed by Karlsson, *J. Immunol. Methods* 145:229-40, 1991 and Cunningham and Wells, *J. Mol. Biol.* 234:554-63, 1993. A receptor, antibody, member or fragment is covalently attached, using amine or sulfhydryl chemistry, to dextran fibers that are attached to gold film within the flow cell. A test sample is passed through the cell. If a ligand, epitope, or opposite member of the complement/anti-complement pair is present in the sample, it will bind to the immobilized receptor, antibody or member, respectively, causing a change in the refractive index of the medium, which is detected as a change in surface plasmon resonance of the gold film. This system allows the determination of on- and off-rates, from which binding affinity can be calculated, and assessment of stoichiometry of binding. As used herein, the term complement/anti-complement pair denotes non-identical moieties that form a non-covalently associated, stable pair under appropriate conditions. For instance, biotin and avidin (or streptavidin) are prototypical members of a complement/anti-complement pair. Other exemplary complement/anti-complement pairs include receptor/ligand pairs, antibody/antigen (or hapten or epitope) pairs, sense/antisense polynucleotide pairs, and the like. Where subsequent dissociation of the complement/anti-complement pair is desirable, the complement/anti-complement pair preferably has a binding affinity of $<10^9 \text{ M}^{-1}$.

Zpep10 polypeptide and other ligand homologs can also be used within other assay systems known in the art. Such systems include Scatchard analysis for determination of binding affinity (see Scatchard, *Ann. NY Acad. Sci.* 51: 660-72, 1949) and calorimetric assays (Cunningham et al., *Science* 253:545-8, 1991; Cunningham et al., *Science* 245:821-5, 1991).

The invention also provides anti-zpep10 antibodies. Antibodies to zpep10 can be obtained, for example, using as an antigen the product of a zpep10 expression vector, or zpep10 isolated from a natural source. Particularly useful anti-zpep10 antibodies "bind specifically" with zpep10. Antibodies are considered to be specifically binding if the antibodies bind to a zpep10 polypeptide, peptide or epitope with a binding affinity (K_a) of 10^6 M^{-1} or greater, preferably 10^7 M^{-1} or greater, more preferably 10^8 M^{-1} or greater, and most preferably 10^9 M^{-1} or greater. The binding affinity of an antibody can be readily determined by one of ordinary skill in the art, for example, by Scatchard analysis (Scatchard, *Ann. NY Acad. Sci.* 51:660, 1949). Suitable antibodies

include antibodies that bind with zpep10, in particular the extracellular domain of zpep10 (amino acid residues 21-111 of SEQ ID NO:2).

Anti-zpep10 antibodies can be produced using antigenic zpep10 epitope-bearing peptides and polypeptides. Antigenic epitope-bearing peptides and polypeptides of the present invention contain a sequence of at least nine, preferably between 15 to about 30 amino acids contained within SEQ ID NO:2. However, peptides or polypeptides comprising a larger portion of an amino acid sequence of the invention, containing from 30 to 50 amino acids, or any length up to and including the entire amino acid sequence of a polypeptide of the invention, also are useful for inducing antibodies that bind with zpep10. It is desirable that the amino acid sequence of the epitope-bearing peptide is selected to provide substantial solubility in aqueous solvents (i.e., the sequence includes relatively hydrophilic residues, while hydrophobic residues are preferably avoided). The hydrophobicity plot provided in the FIGURE provides such information. Using the plot antigenic regions can be selected, such as those found in the fragments, amino acid residue 39-44, 65-70, 38-43, 62-67 and 96-101 of SEQ ID NO:2. Moreover, amino acid sequences containing proline residues may be also be desirable for antibody production.

Polyclonal antibodies to recombinant zpep10 protein or to zpep10 isolated from natural sources can be prepared using methods well-known to those of skill in the art. See, for example, Green et al., "Production of Polyclonal Antisera," in *Immunochemical Protocols* (Manson, ed.), pages 1-5 (Humana Press 1992), and Williams et al., "Expression of foreign proteins in *E. coli* using plasmid vectors and purification of specific polyclonal antibodies," in *DNA Cloning 2: Expression Systems*, 2nd Edition, Glover et al. (eds.), page 15 (Oxford University Press 1995). The immunogenicity of a zpep10 polypeptide can be increased through the use of an adjuvant, such as alum (aluminum hydroxide) or Freund's complete or incomplete adjuvant. Polypeptides useful for immunization also include fusion polypeptides, such as fusions of zpep10 or a portion thereof with an immunoglobulin polypeptide or with maltose binding protein. The polypeptide immunogen may be a full-length molecule or a portion thereof. If the polypeptide portion is "hapten-like," such portion may be advantageously joined or linked to a macromolecular carrier (such as keyhole limpet hemocyanin (KLH), bovine serum albumin (BSA) or tetanus toxoid) for immunization.

Although polyclonal antibodies are typically raised in animals such as horses, cows, dogs, chicken, rats, mice, rabbits, hamsters, guinea pigs, goats or sheep, an anti-zpep10 antibody of the present invention may also be derived from a subhuman primate antibody. General techniques for raising diagnostically and therapeutically useful antibodies in baboons may be found, for example, in Goldenberg et al., international patent publication No. WO 91/11465, and in Losman et al., *Int. J. Cancer* 46:310, 1990. Antibodies can also be raised in transgenic animals such as transgenic sheep, cows, goats or pigs, and may be expressed in yeast and fungi in modified forms as well as in mammalian and insect cells.

Alternatively, monoclonal anti-zpep10 antibodies can be generated. Rodent monoclonal antibodies to specific antigens may be obtained by methods known to those skilled in the art (see, for example, Kohler et al., *Nature* 256:495 (1975), Coligan et al. (eds.), *Current Protocols in Immunology*, Vol. 1, pages 2.5.1-2.6.7 (John Wiley & Sons 1991), Picklesley et al., "Production of monoclonal antibodies against proteins expressed in *E. coli*," in *DNA Cloning 2:*

Expression Systems, 2nd Edition, Glover et al. (eds.), page 93 (Oxford University Press 1995)).

Briefly, monoclonal antibodies can be obtained by injecting mice with a composition comprising a zpep10 gene product, verifying the presence of antibody production by removing a serum sample, removing the spleen to obtain B-lymphocytes, fusing the B-lymphocytes with myeloma cells to produce hybridomas, cloning the hybridomas, selecting positive clones which produce antibodies to the antigen, culturing the clones that produce antibodies to the antigen, and isolating the antibodies from the hybridoma cultures.

In addition, an anti-zpep10 antibody of the present invention may be derived from a human monoclonal antibody. Human monoclonal antibodies are obtained from transgenic mice that have been engineered to produce specific human antibodies in response to antigenic challenge. In this technique, elements of the human heavy and light chain locus are introduced into strains of mice derived from embryonic stem cell lines that contain targeted disruptions of the endogenous heavy chain and light chain loci. The transgenic mice can synthesize human antibodies specific for human antigens, and the mice can be used to produce human antibody-secreting hybridomas. Methods for obtaining human antibodies from transgenic mice are described, for example, by Green et al., *Nat. Genet.* 7:13, 1994, Lonberg et al., *Nature* 368:856, 1994, and Taylor et al., *Int. Immun.* 6:579, 1994.

Monoclonal antibodies can be isolated and purified from hybridoma cultures by a variety of well-established techniques. Such isolation techniques include affinity chromatography with Protein-A Sepharose, size-exclusion chromatography, and ion-exchange chromatography (see, for example, Coligan at pages 2.7.1-2.7.12 and pages 2.9.1-2.9.3; Baines et al., "Purification of Immunoglobulin G (IgG)," in *Methods in Molecular Biology*, Vol. 10, pages 79-104 (The Humana Press, Inc. 1992)).

For particular uses, it may be desirable to prepare fragments of anti-zpep10 antibodies. Such antibody fragments can be obtained, for example, by proteolytic hydrolysis of the antibody. Antibody fragments can be obtained by pepsin or papain digestion of whole antibodies by conventional methods. As an illustration, antibody fragments can be produced by enzymatic cleavage of antibodies with pepsin to provide a 5S fragment denoted F(ab')₂. This fragment can be further cleaved using a thiol reducing agent to produce 3.5 S Fab' monovalent fragments. Optionally, the cleavage reaction can be performed using a blocking group for the sulfhydryl groups that result from cleavage of disulfide linkages. As an alternative, an enzymatic cleavage using pepsin produces two monovalent Fab fragments and an Fc fragment directly. These methods are described, for example, by Goldenberg, U.S. Pat. No. 4,331,647, Nisonoff et al., *Arch Biochem. Biophys.* 89:230, 1960, Porter, *Biochem. J.* 73:119, 1959, Edelman et al., in *Methods in Enzymology* Vol. 1, page 422 (Academic Press 1967), and by Coligan, *ibid*.

Other methods of cleaving antibodies, such as separation of heavy chains to form monovalent light-heavy chain fragments, further cleavage of fragments, or other enzymatic, chemical or genetic techniques may also be used, so long as the fragments bind to the antigen that is recognized by the intact antibody.

For example, Fv fragments comprise an association of V_H and V_L chains. This association can be noncovalent, as described by Inbar et al., *Proc. Nat'l Acad. Sci. USA* 69:2659, 1972. Alternatively, the variable chains can be linked by an intermolecular disulfide bond or cross-linked

by chemicals such as glutaraldehyde (see, for example, Sandhu, *Crit. Rev. Biotech.* 12:437, 1992).

The Fv fragments may comprise V_H and V_L chains which are connected by a peptide linker. These single-chain antigen binding proteins (scFv) are prepared by constructing a structural gene comprising DNA sequences encoding the V_H and V_L domains which are connected by an oligonucleotide. The structural gene is inserted into an expression vector which is subsequently introduced into a host cell, such as *E. coli*. The recombinant host cells synthesize a single polypeptide chain with a linker peptide bridging the two V domains. Methods for producing scFvs are described, for example, by Whitlow et al., *Methods: A Companion to Methods in Enzymology* 2:97, 1991, also see, Bird et al., *Science* 242:423, 1988, Ladner et al., U.S. Pat. No. 4,946,778, Pack et al., *Bio/Technology* 11:1271, 1993, and Sandhu, *supra*.

As an illustration, a scFV can be obtained by exposing lymphocytes to zpep10 polypeptide in vitro, and selecting antibody display libraries in phage or similar vectors (for instance, through use of immobilized or labeled zpep10 protein or peptide). Genes encoding polypeptides having potential zpep10 polypeptide binding domains can be obtained by screening random peptide libraries displayed on phage (phage display) or on bacteria, such as *E. coli*. Nucleotide sequences encoding the polypeptides can be obtained in a number of ways, such as through random mutagenesis and random polynucleotide synthesis. These random peptide display libraries can be used to screen for peptides which interact with a known target which can be a protein or polypeptide, such as a ligand or receptor, a biological or synthetic macromolecule, or organic or inorganic substances. Techniques for creating and screening such random peptide display libraries are known in the art (Ladner et al., U.S. Pat. No. 5,223,409, Ladner et al., U.S. Pat. No. 4,946,778, Ladner et al., U.S. Pat. No. 5,403,484, Ladner et al., U.S. Pat. No. 5,571,698, and Kay et al., *Phage Display of Peptides and Proteins* (Academic Press, Inc. 1996)) and random peptide display libraries and kits for screening such libraries are available commercially, for instance from Clontech (Palo Alto, Calif.), Invitrogen Inc. (San Diego, Calif.), New England Biolabs, Inc. (Beverly, Mass.), and Pharmacia LKB Biotechnology Inc. (Piscataway, N.J.). Random peptide display libraries can be screened using the zpep10 sequences disclosed herein to identify proteins which bind to zpep10.

Another form of an antibody fragment is a peptide coding for a single complementarity-determining region (CDR). CDR peptides ("minimal recognition units") can be obtained by constructing genes encoding the CDR of an antibody of interest. Such genes are prepared, for example, by using the polymerase chain reaction to synthesize the variable region from RNA of antibody-producing cells (see, for example, Larrick et al., *Methods: A Companion to Methods in Enzymology* 2:106, 1991), Courtenay-Luck, "Genetic Manipulation of Monoclonal Antibodies," in *Monoclonal Antibodies: Production, Engineering and Clinical Application*, Ritter et al. (eds.), page 166 (Cambridge University Press 1995), and Ward et al., "Genetic Manipulation and Expression of Antibodies," in *Monoclonal Antibodies: Principles and Applications*, Birch et al., (eds.), page 137 (Wiley-Liss, Inc. 1995)).

Alternatively, an anti-zpep10 antibody may be derived from a "humanized" monoclonal antibody. Humanized monoclonal antibodies are produced by transferring mouse complementary determining regions from heavy and light variable chains of the mouse immunoglobulin into a human variable domain. Typical residues of human antibodies are

then substituted in the framework regions of the murine counterparts. The use of antibody components derived from humanized monoclonal antibodies obviates potential problems associated with the immunogenicity of murine constant regions. General techniques for cloning murine immunoglobulin variable domains are described, for example, by Orlandi et al., *Proc. Natl. Acad. Sci. USA* 86:3833, 1989. Techniques for producing humanized monoclonal antibodies are described, for example, by Jones et al., *Nature* 321:522, 1986, Carter et al., *Proc. Natl. Acad. Sci. USA* 89:4285, 1992, Sandhu, *Crit. Rev. Biotech.* 12:437, 1992, Singer et al., *J. Immunol.* 150:2844, 1993, Sudhir (ed.), *Antibody Engineering Protocols* (Humana Press, Inc. 1995), Kelley, "Engineering Therapeutic Antibodies," in *Protein Engineering: Principles and Practice*, Cleland et al. (eds.), pages 399-434 (John Wiley & Sons, Inc. 1996), and by Queen et al., U.S. Pat. No. 5,693,762 (1997).

Polyclonal anti-idiotypic antibodies can be prepared by immunizing animals with anti-zp10 antibodies or antibody fragments, using standard techniques. See, for example, Green et al., "Production of Polyclonal Antisera," in *Methods In Molecular Biology: Immunochemical Protocols*, Manson (ed.), pages 1-12 (Humana Press 1992). Also, see Coligan, *ibid.* at pages 2.4.1-2.4.7. Alternatively, monoclonal anti-idiotypic antibodies can be prepared using anti-zp10 antibodies or antibody fragments as immunogens with the techniques, described above. As another alternative, humanized anti-idiotypic antibodies or subhuman primate anti-idiotypic antibodies can be prepared using the above-described techniques. Methods for producing anti-idiotypic antibodies are described, for example, by Irie, U.S. Pat. No. 5,208,146, Greene, et. al., U.S. Pat. No. 5,637,677, and Varthakavi and Minocha, *J. Gen. Virol.* 77:1875, 1996.

Antibodies or polypeptides herein can also be directly or indirectly conjugated to drugs, toxins, radionuclides and the like, and these conjugates used for in vivo diagnostic or therapeutic applications. For instance, polypeptides or antibodies of the present invention can be used to identify or treat tissues or organs that express a corresponding anti-complementary molecule (receptor or antigen, respectively, for instance). More specifically, zp10 polypeptides or anti-zp10 antibodies, or bioactive fragments or portions thereof, can be coupled to detectable or cytotoxic molecules and delivered to a mammal having cells, tissues or organs that express the anti-complementary molecule.

Suitable detectable molecules may be directly or indirectly attached to the polypeptide or antibody, and include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent markers, chemiluminescent markers, magnetic particles and the like. Suitable cytotoxic molecules may be directly or indirectly attached to the polypeptide or antibody, and include bacterial or plant toxins (for instance, diphtheria toxin, *Pseudomonas* exotoxin, ricin, abrin and the like), as well as therapeutic radionuclides, such as iodine-131, rhenium-188 or yttrium-90 (either directly attached to the polypeptide or antibody, or indirectly attached through means of a chelating moiety, for instance). Polypeptides or antibodies may also be conjugated to cytotoxic drugs, such as adriamycin. For indirect attachment of a detectable or cytotoxic molecule, the detectable or cytotoxic molecule can be conjugated with a member of a complementary/anticomplementary pair, where the other member is bound to the polypeptide or antibody portion. For these purposes, biotin/streptavidin is an exemplary complementary/anticomplementary pair.

Soluble zp10 polypeptides or antibodies to zp10 can be directly or indirectly conjugated to drugs, toxins, radio-

nuclides and the like, and these conjugates used for in vivo diagnostic or therapeutic applications. For instance, polypeptides or antibodies of the present invention can be used to identify or treat tissues or organs that express a corresponding anti-complementary molecule (receptor or antigen, respectively, for instance). More specifically, zp10 polypeptides or anti-zp10 antibodies, or bioactive fragments or portions thereof, can be coupled to detectable or cytotoxic molecules and delivered to a mammal having cells, tissues or organs that express the anti-complementary molecule.

Suitable detectable molecules can be directly or indirectly attached to the polypeptide or antibody, and include radionuclides, enzymes, substrates, cofactors, inhibitors, fluorescent markers, chemiluminescent markers, magnetic particles and the like. Suitable cytotoxic molecules can be directly or indirectly attached to the polypeptide or antibody, and include bacterial or plant toxins (for instance, diphtheria toxin, *Pseudomonas* exotoxin, ricin, abrin and the like), as well as therapeutic radionuclides, such as iodine-131, rhenium-188 or yttrium-90 (either directly attached to the polypeptide or antibody, or indirectly attached through means of a chelating moiety, for instance). Polypeptides or antibodies can also be conjugated to cytotoxic drugs, such as adriamycin. For indirect attachment of a detectable or cytotoxic molecule, the detectable or cytotoxic molecule can be conjugated with a member of a complementary/anticomplementary pair, where the other member is bound to the polypeptide or antibody portion. For these purposes, biotin/streptavidin is an exemplary complementary/anticomplementary pair.

Such polypeptide-toxin fusion proteins or antibody/fragment-toxin fusion proteins can be used for targeted cell or tissue inhibition or ablation (for instance, to treat cancer cells or tissues). Alternatively, if the polypeptide has multiple functional domains (i.e., an activation domain or a ligand binding domain, plus a targeting domain), a fusion protein including only the targeting domain can be suitable for directing a detectable molecule, a cytotoxic molecule or a complementary molecule to a cell or tissue type of interest. In instances where the domain only fusion protein includes a complementary molecule, the anti-complementary molecule can be conjugated to a detectable or cytotoxic molecule. Such domain-complementary molecule fusion proteins thus represent a generic targeting vehicle for cell/tissue-specific delivery of generic anti-complementary-detectable/cytotoxic molecule conjugates. The bioactive polypeptide or antibody conjugates described herein can be delivered intravenously, intraarterially or intraductally, or may be introduced locally at the intended site of action.

The zp10 gene is almost exclusively expressed in the testis. Low levels of transcript are also seen in a number of other tissues, with the kidney accounting for most of the ancillary expression. The tissue specificity observed for zp10 suggests a general role in development and regulatory control of testicular differentiation and gonadal steroidogenesis and spermatogenesis. Zp10 polypeptides, agonists and antagonists have enormous potential in both in vitro and in vivo applications.

Development of testicular hormone production can be divided into early and late steps, with the latter dependent on the activation of functionally-determined Leydig cell precursors by LH. However, the factors that control the early steps in this process remain unknown (Huhtaniemi, *Reprod. Fertil. Dev.* 7: 1025-35, 1995) suggesting that testis specific polypeptides such as zp10 might be responsible for activation of a non-steroidogenic, non-LH responsive precursor cell.

Once Leydig cell differentiation has occurred, production of steroid hormones in the testis is dependent on the secretion of the gonadotrophins, LH and FSH, by the pituitary. LH stimulates production of testosterone by the Leydig cells, whereas spermatogenesis depends on both FSH and high intratesticular testosterone concentrations. LH and FSH secretion is in turn under control of gonadotrophin releasing hormone (GnRH) produced in the hypothalamus (Kaufman, *The neuro endocrine regulation of male reproduction*. in: Male Infertility. Clinical Investigation, Cause Evaluation and Treatment., F H Comhaire, ed., Chapman and Hall, London, pp 29-54, 1996). Since testicular products have been shown to control LH and FSH production and in turn, these products regulate, testicular function, this suggests a regulatory role for zpep10 in hormone production by the hypothalamic, pituitary, gonadal axis.

It is well known that steroidogenesis and spermatogenesis take place within two different cellular compartments of the testes, with Leydig and Sertoli cells responsible for the former and latter, respectively (Saez, *Endocrin. Rev.* 15: 574-626, 1994). The activity of each of these cell types appears to be regulated by the secretory products of the other. Sertoli cell derived tumor necrosis factor- α , fibroblast growth factor, interleukin-1 transforming growth factor- β , epidermal growth factor/transforming growth factor- α , activin, inhibin, insulin-like growth factor-1, platelet derived growth factor, endothelin, and arginine-vasopressin have all been shown to regulate Leydig cell function (Saez, *Endocrin. Rev.* 15: 574-626, 1994). Thus, zpep10 might control or modulate the activities of one or more of these genes.

The membrane glycoprotein zpep10 may also function as a binding site for one or more growth factor peptides or hormones in much the same way that heparin binds with platelet-derived growth factor (PDGF), fibroblast growth factors (such as aFGF and bFGF) and vascular endothelial growth factor (VEGF) and sequesters them on the cell surface.

In men, aging is associated with a progressive decline in testicular function. These changes are manifest clinically by decreased virility, vigor, and libido that point towards a relative testicular deficiency (Vermeulen, *Ann. Med.* 25:531-4, 1993; Pugeat et al., *Horm. Res.* 43: 104-10, 1995). Hormone replacement therapy in elderly men is not currently recommended which suggests that a new therapy for the male climacterium would be very valuable. Zpep10 polypeptides, agonists or antagonists, either independently or in combination with other factors, may be evaluated therapeutically.

Soluble zpep10 polypeptides, zpep10 agonists and/or zpep10 antagonists may also have therapeutic value in treatment of testicular cancer, infertility, or in the recovery of function following testicular surgery.

The ability of zpep10 polypeptides and zpep10 agonists to stimulate proliferation or differentiation of testicular cells can be measured using cultured testicular cells or in vivo by administering molecules of the present invention to the appropriate animal model. Cultured testicular cells include dolphin DB1.1es cells (CRL-6258); mouse GC-1 spg cells (CRL-2053); TM3 cells (CRL-1714); TM4 cells (CRL-1715); and pig ST cells (CRL-1746), available from American Type Culture Collection, 10801 University Boulevard, Manassas, Va. Assays measuring cell proliferation or differentiation are well known in the art. For example, assays measuring proliferation include such assays as chemosensitivity to neutral red dye (Cavanaugh et al., *Investigational New Drugs* 8:347-354, 1990, incorporation of radiolabelled nucleotides (Cook et al., *Anal. Biochem.* 179:1-7, 1989),

incorporation of 5-bromo-2'-deoxyuridine (BrdU) in the DNA of proliferating cells (Porstmann et al., *J. Immunol. Methods* 82:169-79, 1985), and use of tetrazolium salts (Mosmann, *J. Immunol. Methods* 65:55-63, 1983; Alley et al., *Cancer Res.* 48:589-601, 1988; Marshall et al., *Growth Reg.* 5:69-84, 1995; and Scudiero et al., *Cancer Res.* 48:4827-33, 1988) and by measuring proliferation using ^3H -thymidine uptake (Crowley et al., *J. Immunol. Meth.* 133:55-66, 1990). Assays measuring differentiation include, for example, measuring cell-surface markers associated with stage-specific expression of a tissue, enzymatic activity, functional activity or morphological changes (Watt, *FASEB*, 5:281-4, 1991; Francis, *Differentiation* 57:63-75, 1994; Raes, *Adv. Anim. Cell Biol. Technol. Bioprocesses*, 161-71, 1989).

Zpep10 polypeptides, agonists and antagonists will also prove useful in the study of spermatogenesis and infertility. In vivo, zpep10 agonists may find application in the treatment of male infertility. Zpep10 antagonists may be useful as male contraceptive agents. Zpep10 antagonists are useful as research reagents for characterizing sites of ligand-receptor interaction.

In vivo assays, well known in the art, are available for evaluating the effect of zpep10 ligands and agonists on testes. For example, compounds can be injected intraperitoneally for a specific time duration. After the treatment period, animals are sacrificed and testes removed and weighed. Testicles are homogenized and sperm head counts are made (Meistrich et al., *Exp. Cell Res.* 99:72-8, 1976). Other activities, for example, chemotactic activity that may be associated with proteins of the present invention can be analyzed. For example, late stage factors in spermatogenesis are involved in egg-sperm interactions and sperm motility. Activities, such as enhancing viability of cryopreserved sperm, stimulating the acrosome reaction, enhancing sperm motility and enhancing egg-sperm interactions may be associated with the ligands and agonists of the present invention. Assays evaluating such activities are known (Rosenberger, *J. Androl.* 11:89-96, 1990; Fuchs, *Zentralbl. Gynakol.* 11:117-120, 1993; Neurwinger et al., *Andrologia* 22:335-9, 1990; Harris et al., *Human Reprod.* 3:856-60, 1988; and Jockenhovel, *Andrologia* 22:171-178, 1990; Lessing et al., *Fertil. Steril.* 44:406-9, 1985; Zaneveld, In Male Infertility Chapter 11, Comhaire Ed., Chapman & Hall, London 1996). These activities are expected to result in enhanced fertility and successful reproduction.

Localization of zpep10 to testis tissue suggests zpep10, its agonists and/or antagonists may have applications in enhancing fertilization during assisted reproduction in humans and in animals. Such assisted reproduction methods are known in the art and include artificial insemination, in vitro fertilization, embryo transfer and gamete intrafallopian transfer. Such methods are useful for assisting men and women who may have physiological or metabolic disorders that prevent natural conception. Such methods are also used in animal breeding programs, such as for livestock, zoological animals, endangered species or racehorses and could be used as methods for the creation of transgenic animals.

To verify the presence of this capability in zpep10 polypeptides, agonists or antagonists of the present invention, such molecules are evaluated with respect to their ability to enhance viability of cryopreserved sperm, sperm motility, the ability of sperm to penetrate cervical mucus, particularly in association with methods of assisted reproduction, according to procedures known in the art (see for example, Juang et al., *Anim. Reprod. Sci.* 20:21-9, 1989; Juang et al., *Anim. Reprod. Sci.* 22:47-53, 1990; Colon et al.,

Fertil. Steril. 46:1133-39, 1986; Lessing et al., *Fertil. Steril.* 44:406-9, 1985 and Brenner et al., *Fertil. Steril.* 42:92-6, 1984). If desired, zpep10 polypeptide performance in this regard can be compared to relaxins and the like. In addition, zpep10 polypeptides or agonists or antagonists thereof may be evaluated in combination with one or more proteins to identify synergistic effects. For example, soluble zpep10, agonists and/or antagonists can be added to "capacitation media", a cocktail of compounds known to activate sperm, such as caffeine, dibutyl cyclic adenosine monophosphate (dbcAMP) or theophylline. Such mixtures have resulted in improved reproductive function of the sperm, in particular, sperm motility and zonae penetration (Park et al., *Am. J. Obstet. Gynecol.* 158:974-9, 1988; Vandevoort et al., *Mol. Reprod. Develop.* 37:299-304, 1993; Vandevoort and Overstreet, *J. Androl.* 16:327-33, 1995). The capacitation mixture can then be combined with sperm, an egg or an egg-sperm mixture prior to fertilization of the egg.

In cases where pregnancy is not desired, zpep10 polypeptides or polypeptide fragments may function as germ-cell-specific antigens for use as components in "immunocontraceptive" or "anti-fertility" vaccines to induce formation of antibodies and/or cell mediated immunity to selectively inhibit a process, or processes, critical to successful reproduction in humans and animals. The use of sperm and testis antigens in the development of an immunocontraceptive have been described (O'Hern et al., *Biol. Reprod.* 52:311-39, 1995; Diekman and Herr, *Am. J. Reprod. Immunol.* 37:111-17, 1997; Zhu and Naz, *Proc. Natl. Acad. Sci. USA* 94:4704-9, 1997). A vaccine based on human chorionic gonadotrophin (HCG) linked to a diphtheria or tetanus carrier is currently in clinical trials (Talwar et al., *Proc. Natl. Acad. Sci. USA* 91:8532-36, 1994). A single injection resulted in production of high titer antibodies that persisted for nearly a year in rabbits (Stevens, *Am. J. Reprod. Immunol.* 29:176-88, 1993). Such methods of immunocontraception using vaccines would include a zpep10 testes-specific protein or fragment thereof. The zpep10 protein or fragments can be conjugated to a carrier protein or peptide, such as tetanus or diphtheria toxoid. An adjuvant, as described above, can be included and the protein or fragment can be noncovalently associated with other molecules to enhance intrinsic immunoreactivity. Methods for administration and methods for determining the number of administrations are known in the art. Such a method might include a number of primary injections over several weeks followed by booster injections as needed to maintain a suitable antibody titer.

For pharmaceutical use, pharmaceutically effective amounts of zpep10 therapeutic antibodies, small molecule antagonists or agonists of zpep10 polypeptides, or zpep10 polypeptide fragments or soluble zpep10 receptors can be formulated with pharmaceutically acceptable carriers for parenteral, oral, nasal, rectal, topical, transdermal administration or the like, according to conventional methods. Formulations may further include one or more diluents, fillers, emulsifiers, preservatives, buffers, excipients, and the like, and may be provided in such forms as liquids, powders, emulsions, suppositories, liposomes, transdermal patches and tablets, for example. Slow or extended-release delivery systems, including any of a number of biopolymers (biological-based systems), systems employing liposomes, and polymeric delivery systems, can also be utilized with the compositions described herein to provide a continuous or long-term source of the zpep10 polypeptide, agonist or antagonist. Such slow release systems are applicable to formulations, for example, for oral, topical and parenteral

use. The term "pharmaceutically acceptable carrier or vehicle" refers to a carrier medium which does not interfere with the effectiveness of the biological activity of the active ingredients and which is not toxic to the host or patient. One skilled in the art may formulate the compounds of the present invention in an appropriate manner, and in accordance with accepted practices, such as those disclosed in Remington: *The Science and Practice of Pharmacy*, Gennaro, ed., Mack Publishing Co., Easton, Pa., 19th ed., 1995.

As used herein, a pharmaceutically effective amount of a zpep10 polypeptide, agonist or antagonist, is an amount sufficient to induce a desired biological result. The result can be alleviation of the signs, symptoms, or causes of a disease, or any other desired alteration of a biological system. For example, an effective amount of a polypeptide of the present invention is that which provides either subjective relief of symptoms or an objectively identifiable improvement as noted by the clinician or other qualified observer. Doses of zpep10 polypeptide will generally be determined by the clinician according to accepted standards, taking into account the nature and severity of the condition to be treated, patient traits, etc. Determination of dose is within the level of ordinary skill in the art. The proteins may be administered for acute treatment, over one week or less, often over a period of one to three days or may be used in chronic treatment, over several months or years.

Radiation hybrid mapping is a somatic cell genetic technique developed for constructing high-resolution, contiguous maps of mammalian chromosomes (Cox et al., *Science* 250:245-50, 1990). Partial or full knowledge of a gene's sequence allows the designing of PCR primers suitable for use with chromosomal radiation hybrid mapping panels. Radiation hybrid mapping panels are commercially available which cover the entire human genome, such as the Stanford G3 RH Panel and the GeneBridge 4 RH Panel (Research Genetics, Inc., Huntsville, Ala.). These panels enable rapid, PCR based, chromosomal localizations and ordering of genes, sequence-tagged sites (STSs), and other nonpolymorphic and polymorphic markers within a region of interest. This includes establishing directly proportional physical distances between newly discovered genes of interest and previously mapped markers. The precise knowledge of a gene's position can be useful in a number of ways including: 1) determining if a sequence is part of an existing contig and obtaining additional surrounding genetic sequences in various forms such as YAC-, BAC- or cDNA clones, 2) providing a possible candidate gene for an inheritable disease which shows linkage to the same chromosomal region, and 3) for cross-referencing model organisms such as mouse which may be beneficial in helping to determine what function a particular gene might have.

The present invention provides reagents for use in diagnostic applications. For example, the zpep10 gene, a probe comprising zpep10 DNA or RNA, or a subsequence thereof can be used to determine if the zpep10 gene is present on a particular chromosome or if a mutation has occurred. Detectable chromosomal aberrations at the zpep10 gene locus include, but are not limited to, aneuploidy, gene copy number changes, insertions, deletions, restriction site changes and rearrangements. These aberrations can occur within the coding sequence, within introns, or within flanking sequences, including upstream promoter and regulatory regions, and may be manifested as physical alterations within a coding sequence or changes in gene expression level.

In general, these diagnostic methods comprise the steps of (a) obtaining a genetic sample from a patient; (b) incubating

the genetic sample with a polynucleotide probe or primer as disclosed above, under conditions wherein the polynucleotide will hybridize to complementary polynucleotide sequence, to produce a first reaction product; and (iii) comparing the first reaction product to a control reaction product. A difference between the first reaction product and the control reaction product is indicative of a genetic abnormality in the patient. Genetic samples for use within the present invention include genomic DNA, cDNA, and RNA. The polynucleotide probe or primer can be RNA or DNA, and will comprise a portion of SEQ ID NO:1, the complement of SEQ ID NO:1, or an RNA equivalent thereof. Suitable assay methods in this regard include molecular genetic techniques known to those in the art, such as restriction fragment length polymorphism (RFLP) analysis, short tandem repeat (STR) analysis employing PCR techniques, ligation chain reaction (Barany, *PCR Methods and Applications* 1:5-16, 1991), ribonuclease protection assays, and other genetic linkage analysis techniques known in the art (Sambrook et al., *ibid.*; Ausubel et al., *ibid.*; Marian, *Chest* 108:255-65, 1995). Ribonuclease protection assays (see, e.g., Ausubel et al., *ibid.*, ch. 4) comprise the hybridization of an RNA probe to a patient RNA sample, after which the reaction product (RNA-RNA hybrid) is exposed to RNase. Hybridized regions of the RNA are protected from digestion. Within PCR assays, a patient's genetic sample is incubated with a pair of polynucleotide primers, and the region between the primers is amplified and recovered. Changes in size or amount of recovered product are indicative of mutations in the patient. Another PCR-based technique that can be employed is single strand conformational polymorphism (SSCP) analysis (Hayashi, *PCR Methods and Applications* 1:34-8, 1991).

Polynucleotides encoding zpep10 polypeptides are useful within gene therapy applications where it is desired to increase or inhibit zpep10 activity. If a mammal has a mutated or absent zpep10 gene, the zpep10 gene can be introduced into the cells of the mammal. In one embodiment, a gene encoding a zpep10 polypeptide is introduced in vivo in a viral vector. Such vectors include an attenuated or defective DNA virus, such as, but not limited to, herpes simplex virus (HSV), papillomavirus, Epstein Barr virus (EBV), adenovirus, adeno-associated virus (AAV), and the like. Defective viruses, which entirely or almost entirely lack viral genes, are preferred. A defective virus is not infective after introduction into a cell. Use of defective viral vectors allows for administration to cells in a specific, localized area, without concern that the vector can infect other cells. Examples of particular vectors include, but are not limited to, a defective herpes simplex virus 1 (HSV1) vector (Kaplitt et al., *Molec. Cell. Neurosci.* 2:320-30, 1991); an attenuated adenovirus vector, such as the vector described by Stratford-Perricaudet et al., *J. Clin. Invest.* 90:626-30, 1992; and a defective adeno-associated virus vector (Samulski et al., *J. Virol.* 61:3096-101, 1987; Samulski et al., *J. Virol.* 63:3822-8, 1989).

In another embodiment, a zpep10 gene can be introduced in a retroviral vector, e.g., as described in Anderson et al., U.S. Pat. No. 5,399,346; Mann et al. *Cell* 33:153, 1983; Temin et al., U.S. Pat. No. 4,650,764; Temin et al., U.S. Pat. No. 4,980,289; Markowitz et al., *J. Virol.* 62:1120, 1988; Temin et al., U.S. Pat. No. 5,124,263; International Patent Publication No. WO 95/07358, published Mar. 16, 1995 by Dougherty et al.; and Kuo et al., *Blood* 82:845, 1993. Alternatively, the vector can be introduced by lipofection in vivo using liposomes. Synthetic cationic lipids can be used to prepare liposomes for in vivo transfection of a gene

encoding a marker (Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7, 1987; Mackey et al., *Proc. Natl. Acad. Sci. USA* 85:8027-31, 1988). The use of lipofection to introduce exogenous genes into specific organs in vivo has certain practical advantages. Molecular targeting of liposomes to specific cells represents one area of benefit. More particularly, directing transfection to particular cells represents one area of benefit. For instance, directing transfection to particular cell types would be particularly advantageous in a tissue with cellular heterogeneity, such as the pancreas, liver, kidney, and brain. Lipids may be chemically coupled to other molecules for the purpose of targeting. Targeted peptides (e.g., hormones or neurotransmitters), proteins such as antibodies, or non-peptide molecules can be coupled to liposomes chemically.

It is possible to remove the target cells from the body; to introduce the vector as a naked DNA plasmid; and then to re-implant the transformed cells into the body. Naked DNA vectors for gene therapy can be introduced into the desired host cells by methods known in the art, e.g., transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, use of a gene gun or use of a DNA vector transporter. See, e.g., Wu et al., *J. Biol. Chem.* 267:963-7, 1992; Wu et al., *J. Biol. Chem.* 263:14621-4, 1988.

Antisense methodology can be used to inhibit zpep10 gene translation, such as to inhibit cell proliferation in vivo. Polynucleotides that are complementary to a segment of a zpep10-encoding polynucleotide (e.g., a polynucleotide as set forth in SEQ ID NO:1) are designed to bind to zpep10-encoding mRNA and to inhibit translation of such mRNA. Such antisense polynucleotides are used to inhibit expression of zpep10 polypeptide-encoding genes in cell culture or in a subject.

Transgenic mice, engineered to express the zpep10 gene, and mice that exhibit a complete absence of zpep10 gene function, referred to as "knockout mice" (Snouwaert et al., *Science* 257:1083, 1992), may also be generated (Lowell et al., *Nature* 366:740-42, 1993). These mice may be employed to study the zpep10 gene and the protein encoded thereby in an in vivo system.

The invention is further illustrated by the following non-limiting examples.

EXAMPLES

Example 1

Identification of Zpep10

The zpep10 polypeptide-encoding polynucleotides of the present invention were initially identified by querying an EST database for polypeptides containing repetitive patterns and post-translational processing sites yielding potentially active peptides. The polypeptide encoded by an EST meeting those search criteria was further analyzed and found to be a membrane glycoprotein. The EST sequence was from a testis cell library. Several clones considered likely to contain the entire coding region were used for sequencing and resulted in incompletely spliced messages. A minimal nucleotide sequence having all potential introns spliced out was generated from these sequences.

To obtain the complete cDNA sequence a human testis library was screened. The library was plated in pools of 12,000. Plasmid DNA was prepared from the plated bacteria using a Qiagen® plasmid purification column (Qiagen, Inc., Chatsworth, Calif.) according to the manufacturer's instructions. DNA from these pools were used as template DNA to

identify pools containing the DNA encoding zpep10 using PCR. Oligonucleotide primers ZC16,186, (SEQ ID NO:9) and ZC16,187, (SEQ ID NO:10) were designed from the sequence of the EST. One nanogram of template DNA was combined with 20 pmoles of each primer in a PCR mixture. The reaction mixture was incubated at 94° C. for 5 minutes, then run for 35 cycles of 94° C., 30 seconds and 68° C., 30 seconds; followed by an extension at 68° C. for 7 minutes. Pools having the correct sized PCR produce, 290 bp, were used as a template for PCR isolation of the 5' end of the clones. Sequence specific primer ZC16,186 (SEQ ID NO:9) and vector specific primer ZC13,006 (SEQ ID NO:11) were used in PCR reactions as above. PCR products were purified by Qiaex II Gel Extraction Kit (Qiagen, Inc.) according to manufacturer's instructions and sequenced. Pools which contained the clones with the most fully spliced sequence were used to transform *E. coli* and plated to agar. The colonies were transferred to nitrocellulose and probed with a 290 bp fragment derived above. The probe was radioactively labeled using the MULTIPRIME DNA labeling kit (Amersham, Arlington Heights, Ill.) according to the manufacturer's instructions. The probe was purified using a NUCTRAP push column (Stratagene). ExpressHyb (Clontech) solution was used for prehybridization and as a hybridizing solution for the colony lifts. Hybridization took place at 65° C. for over 12 hours using 1.2×10⁶ cpm/ml of labeled probe. The filters were then washed 4 times at 5 minutes each in 2× SSC, 0.005% SDS at 25° C. followed by 2 washes at 20 minutes each in 0.1× SSC, 0.1% SDS at 50° C. with continuous agitation. Plasmid DNA from those colonies producing a signal were isolated and sequenced. The 1094 bp (SEQ ID NO:1) sequence encoding the a full

length zpep10 polypeptide was isolated. An intron may be contained within the 3' untranslated region from base pairs 560–784 of SEQ ID NO:1.

Example 2

Tissue Distribution

Human Multiple Tissue Northern Blots (MTN I, MTN II and MTN III; Clontech) were probed to determine the tissue distribution of human zpep10 expression. An approximately 530 bp probe, entirely 3' UTR was derived by restriction digest of the clone described above with Not I and Eco RI. The restriction digested fragment was visualized by agarose gel electrophoresis and purified using Qiaex II (Qiagen, Chatsworth, Calif.) according to manufacturer's instructions. The probe was radioactively labeled using the MULTIPRIME DNA labeling kit (Amersham, Arlington Heights, Ill.) according to the manufacturer's instructions. The probe was purified using a NUCTRAP push column (Stratagene) EXPRESSHYB (Clontech) solution was used for prehybridization and as a hybridizing solution for the Northern blots. Hybridization took place overnight at 65° C. using and the blots were then washed at 50° C. in 1× SSC, 0.1% SDS. A 1.5 kb transcript corresponding to zpep10 was seen in testis and a non-discrete smear was seen in kidney.

A RNA Master Dot Blot (Clontech) that contained RNAs from various tissues that were normalized to 8 housekeeping genes was also probed and hybridized as described above. The highest level of expression was seen in testis with significantly reduced expression in kidney.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 14

<210> SEQ ID NO 1
<211> LENGTH: 1094
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (79)...(504)

<400> SEQUENCE: 1

```

ggcaaggcgt ggagccaggg ctgcagagca ttccttggt cagctggggc agcgccgccc      60
catcccccag tggctctc atg tgg agg ctg gca cta ggc ggg gtt ttc ctg      111
          Met Trp Arg Leu Ala Leu Gly Gly Val Phe Leu
          1              5              10

gca gcc gcc cag gct tgt gtc ttc tgt cgc ctc cca gcc cac gac ttg      159
Ala Ala Ala Gln Ala Cys Val Phe Cys Arg Leu Pro Ala His Asp Leu
          15              20              25

tca ggc cgc ctg gct cgg ctc tgc agc cag atg gag gcc agg cag aag      207
Ser Gly Arg Leu Ala Arg Leu Cys Ser Gln Met Glu Ala Arg Gln Lys
          30              35              40

gaa tgt ggg gcc tcc cca gac ttc tcg gcc ttt gcc tta gat gag gtg      255
Glu Cys Gly Ala Ser Pro Asp Phe Ser Ala Phe Ala Leu Asp Glu Val
          45              50              55

tcc atg aac aaa gtc aca gag aag act cac aga gtc ctg agg gtc atg      303
Ser Met Asn Lys Val Thr Glu Lys Thr His Arg Val Leu Arg Val Met
          60              65              70              75

ggg ggc agc acc acg ctg tac aac tgc tcc acc tgc aag ggg acg gag      351

```

-continued

Gly	Gly	Ser	Thr	Thr	Leu	Tyr	Asn	Cys	Ser	Thr	Cys	Lys	Gly	Thr	Glu	
				80					85					90		
gtg	tcc	tcg	tcg	ccc	cga	aag	cgc	tgc	ttc	cca	gga	agt	cag	gat	ctt	399
Val	Ser	Cys	Trp	Pro	Arg	Lys	Arg	Cys	Phe	Pro	Gly	Ser	Gln	Asp	Leu	
			95					100					105			
tcg	gaa	gcc	aag	att	ctg	ctc	ctc	tcc	atc	ttc	gga	gct	ttc	ctg	ctt	447
Trp	Glu	Ala	Lys	Ile	Leu	Leu	Leu	Ser	Ile	Phe	Gly	Ala	Phe	Leu	Leu	
			110					115					120			
ctg	ggt	gtt	ctg	agc	ctc	ctg	gtg	gag	tcc	cac	cac	ctc	caa	gca	aaa	495
Leu	Gly	Val	Leu	Ser	Leu	Leu	Val	Glu	Ser	His	His	Leu	Gln	Ala	Lys	
			125					130					135			
agt	ggc	ttg	tgaagacgct	gaaaacctcc	cagcctccag	ctctaagggg										544
Ser	Gly	Leu														
			140													
tatgcactca	caacttcac	atcccttga	ggggaaccag	tcagcccctt	agtcccagct											604
ccaaagacag	tctccagacc	ctaaaacca	gacatccctg	cttctgggtg	gtgagataat											664
gaaaaacaag	aaaatcccca	aaaaccaga	tccccacaa	tccagtgct	agatggcctc											724
ccgggaaccc	aggcaccac	agctggaaag	ttctctccct	ccagccctca	accaatcaca											784
tggtgtcaa	caatgccag	aaaatatcta	cagaaggaaa	gaatccccta	cgccactccc											844
accacaccca	caccccttc	tgctgttcc	gggaaagcgg	gggcatctgc	cccagaagct											904
attccaggcc	ctcctatgac	tgatggggaa	tccgggaatg	catgttctgg	aaaactcacc											964
ccactagagt	gagatcacat	cagtgggttc	gcgggcatgc	cctccctcca	tcgtgttaac											1024
agtttgaaat	cctggcctcc	ctcagaggcc	tccatcctgc	caggcctaag	taaaacttgc											1084
tgttcatgga																1094

<210> SEQ ID NO 2

<211> LENGTH: 142

<212> TYPE: PRT

<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 2

Met	Trp	Arg	Leu	Ala	Leu	Gly	Gly	Val	Phe	Leu	Ala	Ala	Ala	Gln	Ala	
1			5					10						15		
Cys	Val	Phe	Cys	Arg	Leu	Pro	Ala	His	Asp	Leu	Ser	Gly	Arg	Leu	Ala	
			20					25					30			
Arg	Leu	Cys	Ser	Gln	Met	Glu	Ala	Arg	Gln	Lys	Glu	Cys	Gly	Ala	Ser	
			35					40					45			
Pro	Asp	Phe	Ser	Ala	Phe	Ala	Leu	Asp	Glu	Val	Ser	Met	Asn	Lys	Val	
			50					55					60			
Thr	Glu	Lys	Thr	His	Arg	Val	Leu	Arg	Val	Met	Gly	Gly	Ser	Thr	Thr	
			65					70					75		80	
Leu	Tyr	Asn	Cys	Ser	Thr	Cys	Lys	Gly	Thr	Glu	Val	Ser	Cys	Trp	Pro	
			85						90					95		
Arg	Lys	Arg	Cys	Phe	Pro	Gly	Ser	Gln	Asp	Leu	Trp	Glu	Ala	Lys	Ile	
			100					105					110			
Leu	Leu	Leu	Ser	Ile	Phe	Gly	Ala	Phe	Leu	Leu	Leu	Gly	Val	Leu	Ser	
			115					120					125			
Leu	Leu	Val	Glu	Ser	His	His	Leu	Gln	Ala	Lys	Ser	Gly	Leu			
			130					135					140			

<210> SEQ ID NO 3

<211> LENGTH: 426

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

-continued

```

<220> FEATURE:
<223> OTHER INFORMATION: Degenerate nucleotide sequence encoding the
      zpep10 polypeptide of SEQ ID NO:2
<221> NAME/KEY: variation
<222> LOCATION: (1)...(426)
<223> OTHER INFORMATION: Each N is independently any nucleotide.

<400> SEQUENCE: 3

atgtggmgny tngcnytnng ngngntntty ytngcngcng cncargcntg ygtnttytgy      60
mgnytnccng cncaygayyt nwsnggmgn ytnncnmgny tntgywsnca ratggargcn      120
mgncaraarg artgyggngc nwsnccngay ttywengcnt tygcnytnga ygargtnwsn      180
atgaayaarg tnacngaraa racncaymgn gtntnmgng tnatggngg nwsnacnacn      240
ytnntayaayt gywsnacntg yaarggnacn gargtnwsnt gytggccnmg naarmgntgy      300
ttyccnggnw sncargayyt ntggggargcn aarathytny tnytnwsnat httyggngcn      360
ttyytnytny tngngntnyt nwsnytnytn gtngarwscn aycayytnca rgcnaarwen      420
ggnytn                                           426

<210> SEQ ID NO 4
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Degenerate oligonucleotide probe
<221> NAME/KEY: variation
<222> LOCATION: (1)...(17)
<223> OTHER INFORMATION: Each N is independently any nucleotide

<400> SEQUENCE: 4

cargcntgyg tnttytg                                           17

<210> SEQ ID NO 5
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Degenerate oligonucleotide probe
<221> NAME/KEY: variation
<222> LOCATION: (1)...(17)
<223> OTHER INFORMATION: Each N is independently any nucleotide.

<400> SEQUENCE: 5

caraargart gyggngc                                           17

<210> SEQ ID NO 6
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Degenerate oligonucleotide probe
<221> NAME/KEY: variation
<222> LOCATION: (1)...(17)
<223> OTHER INFORMATION: Each N is independently any nucleotide.

<400> SEQUENCE: 6

atgaayaarg rncnga                                           17

<210> SEQ ID NO 7
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Degenerate oligonucleotide probe
<221> NAME/KEY: variation
<222> LOCATION: (1)...(17)

```

-continued

<223> OTHER INFORMATION: Each N is independently any nucleotide.

<400> SEQUENCE: 7

grnacngara aracnca

17

<210> SEQ ID NO 8

<211> LENGTH: 17

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Degenerate oligonucleotide probe

<221> NAME/KEY: variation

<222> LOCATION: (1)...(17)

<223> OTHER INFORMATION: Each N is independently any nucleotide.

<400> SEQUENCE: 8

acntgyaarg gnacnga

17

<210> SEQ ID NO 9

<211> LENGTH: 25

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Oligonucleotide ZC16,186

<400> SEQUENCE: 9

atcagtcata ggagggcctg gaata

25

<210> SEQ ID NO 10

<211> LENGTH: 25

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Oligonucleotide ZC16,187

<400> SEQUENCE: 10

tccctgcttc tggttgtga gataa

25

<210> SEQ ID NO 11

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Oligonucleotide ZC 13,006

<400> SEQUENCE: 11

ggctgtcctc taagcgtcac

20

<210> SEQ ID NO 12

<211> LENGTH: 12

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Nucleotide contig example

<400> SEQUENCE: 12

atggcttagc tt

12

<210> SEQ ID NO 13

<211> LENGTH: 12

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Nucleotide contig example

<400> SEQUENCE: 13

-continued

tagcttgagt ct

12

<210> SEQ ID NO 14
 <211> LENGTH: 12
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: Nucleotide contig example

<400> SEQUENCE: 14

gtcgactacc ga

12

15

What is claimed is:

1. An isolated polypeptide comprising an extracellular domain, wherein said extracellular domain comprises amino acid residues 22 to 111 of the amino acid sequence of SEQ ID NO:2.

2. An isolated polypeptide according to claim 1, wherein said polypeptide further comprises a transmembrane domain that resides in a carboxyl-terminal position relative to said extracellular domain, wherein said transmembrane domain comprises amino acid residues 112 to 133 of the amino acid sequence of SEQ ID NO:2.

3. An isolated polypeptide according to claim 2, wherein said polypeptide further comprises a cytoplasmic domain that resides in a carboxyl-terminal position relative to said transmembrane domain, wherein said cytoplasmic domain comprises amino acid residues 134 to 142 of the amino acid sequence of SEQ ID NO:2.

4. An isolated polypeptide according to claim 2, wherein said polypeptide further comprises a secretory signal that resides in an amino-terminal position relative to said extracellular domain, wherein said secretory signal sequence comprises amino acid residues 1 to 20 of the amino acid sequence of SEQ ID NO:2.

5. An isolated polypeptide according to claim 1 comprising amino acid residue 1 to amino acid residue 142 of SEQ ID NO:2.

6. An isolated polypeptide according to claim 1, covalently linked amino terminally or carboxy terminally to a moiety selected from the group consisting of affinity tags, toxins, radionucleotides, enzymes and fluorophores.

7. An isolated polypeptide comprising a sequence of amino acid residues that is at least 80% identical to a amino acid residue 21 to amino acid residue 142 of SEQ ID NO:2, wherein said polypeptide specifically binds with an antibody

that specifically binds with a polypeptide having the amino acid sequence of SEQ ID NO:2.

8. An isolated polypeptide according to claim 7, wherein any difference between said amino acid sequence of said isolated polypeptide and said corresponding amino acid sequence of SEQ ID NO:2 is due to a conservative amino acid substitution.

9. An isolated polypeptide of claim 7, wherein the amino acid percent identity is determined using a FASTA program with ktup=1, gap opening penalty=10, gap extension penalty=1, and substitution matrix=blosum62, with other parameters set as default.

10. An isolated polypeptide comprising the amino acid sequence of amino acid residue 1 to amino acid residue 20 of SEQ ID NO:2.

11. An isolated polypeptide selected from the group consisting of:

- a) amino acid residues 21-111 of SEQ ID NO:2;
- b) amino acid residues 112-133 of SEQ ID NO:2;
- c) amino acid residues 134-142 of SEQ ID NO:2;
- d) amino acid residues 1-20 of SEQ ID NO:2;
- e) amino acid residues 21-133 of SEQ ID NO:2;
- f) amino acid residues 112-142 of SEQ ID NO:2;
- g) amino acid residues 1-111 of SEQ ID NO:2; and
- h) amino acid residues 1-133 of SEQ ID NO:2.

12. A fusion protein consisting of a first portion and a second portion joined by a peptide bond, said first portion comprising a polypeptide according to claim 1, and said second portion comprising another polypeptide.

13. A polypeptide according to claim 1 in combination with a pharmaceutically acceptable vehicle.

* * * * *



US006730502B2

(12) **United States Patent**
Van Hijum et al.

(10) **Patent No.:** **US 6,730,502 B2**
(45) **Date of Patent:** **May 4, 2004**

(54) **FRUCTOSYLTRANSFERASES**

(58) **Field of Search** 435/97, 101, 193,
435/252.9

(75) **Inventors:** **Sacha Adrianus Fokke Taco Van Hijum**, Groningen (NL); **Gerritdina Hendrika Van Geel-Schutten**, Driebergen-Rijsenberg (NL); **Lubbert Dijkhuizen**, Zuidlaren (NL); **Hakim Rahaoui**, Amersfoort (NL)

(56) **References Cited**

PUBLICATIONS

G.H. van Geel-Schutten et al. "Biochemical and Structural Characterization of the Glucan and Fructan Exopolysaccharides Synthesized by the *Lactobacillus reuteri* Wild-Type Strain and by Mutant Strains", *Applied Environ. Microbiol.* 65(7); 3008-3014. (Jul. 1).*

G.H. van Geel-Schutten et al. "Screening and Characterization of *Lactobacillus* Strains Producing Large Amounts of Exopolysaccharides", *Appl. Microbiol. Biotechnol.* 50: 697-703. (1998).*

* cited by examiner

Primary Examiner—Rebecca Prouty

(74) *Attorney, Agent, or Firm*—Young & Thompson

(57) **ABSTRACT**

The present invention describes two novel proteins having fructosyltransferase activity. Both enzymes are derived from lactobacilli, which are food-grade micro-organisms with the Generally Recognized As Safe (GRAS) status. One of these proteins produces an inulin and fructo-oligosaccharides, while the other produces a levan and fructo-oligosaccharides. According to the invention lactobacilli capable of producing an inulin and/or a levan and/or fructo-oligosaccharides using one or both of the fructosyltransferases can be used as a probiotic or a symbiotic.

8 Claims, 8 Drawing Sheets

(73) **Assignee:** **Nederlandse Organisatie voor Toegepast - Natuurwetenschappelijk Onderzoek TNO**, Delft (NL)

(*) **Notice:** Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) **Appl. No.:** **09/995,587**

(22) **Filed:** **Nov. 29, 2001**

(65) **Prior Publication Data**

US 2002/0127681 A1 Sep. 12, 2002

Related U.S. Application Data

(63) Continuation-in-part of application No. 09/604,958, filed on Jun. 28, 2000, now Pat. No. 6,635,460.

(30) **Foreign Application Priority Data**

May 25, 2000 (EP) 00201872

(51) **Int. Cl.⁷** **C12N 9/10**; **C12P 19/18**;
C12P 19/04

(52) **U.S. Cl.** **435/97**; **435/101**; **435/193**;
435/252.9

Fig 1 (1)

1 tacaatgggg tggcggaggt gaagaaacgg ggttacttct atgctagaac gcaaggaaca 19ftf>

y n g v a e v k k r g y f y a r t
y n g v a e v n t e r q a n g q i

61 taaaaaaatg tataaaagcg gtaaaaattg ggcagtcggt acactctcga ctgctgcgct

1 m y k s g k n w a v v t l s t a a

121 ggtatttggt gcaacaactg taaatgcac cycggaacaca aatattgaaa acaatgatcc

18 l v f g a t t v n a s a d t n i e n n d

181 ttctactgta caagttacaa caggtgataa tgatattgct gttaaaagt tgacacttgg

38 s s t v q v t t g d n d i a v k s v t l

241 tagtgggtcaa gttagtgcag ctagtgtac gactattaga acttctgcta atgcaaatag

58 g s g q v s a a s d t t i r t s a n a n

301 tgctttcttct gccgctaata cacaaaattc taacagtcaa gtagcaagtt ctgctgcaat

78 s a s s a a n t q n s n s q v a s s a a

361 aacatcatct acaagttccg cagcttcatt aaataacaca gatagtaaag cggctcaaga

98 i t s s t s s a a s l n n t d s k a a q

421 aaataactaat acagccaaaa atgatgacac gcaaaaagct gcaccagcta acgaatcttc

118 e n t n t a k n d d t q k a a p a n e s

481 tgaagctaaa aatgaaccag ctgtaaactg taatgattct tcagctgcaa aaaatgatga

138 s e a k n e p a v n v n d s s a a k n d

541 tcaacaatcc agtaaaaaga atactaccgc taagttaaac aaggatgctg aaaacgttgt

158 d q q s s k k n t t a k l n k d a e n v

601 aaaaaaggcg ggaattgatc ctaacagttt aactgatgac cagattaaag cattaataa

178 v k k a g i d p n s l t d d q i k a l n

Fig 1 (2)

661 gatgaacttc tcgaaagctg caaagtctgg tacacaaatg acttataatg atttccaaaa
198 k m n f s k a a k s g t q m t y n d f q

721 gattgctgat acgttaatea aacaagatgg tcggtacaca gttccattct ttaaagcaag 20ftfi <
218 k i a d t l i k q d g r y t v p f f k a

781 tgaatcaaaa aatatgcctg ccgtacaaac taaagatgca caaactaata ctattgaacc
238 s e i k n m p a a t t k d a q t n t i e

841 tttagatgta tgggattcat ggccagttca agatgttcgg acaggacaag ttgctaattg 5ftf >
258 p l d v w d s w p v q d v r t g q v a n 8ftfi <

901 gaatggctat caacttgtca tcgcaatgat gggaattcca aaccaaagt ataatcatat
278 w n g y q l v i a m m g i p n q n d n h

961 ctatctctta tataataagt atggtgataa tgaattaagt cattggaaga atgtaggtcc 7ftf >
298 i y l l y n k y g d n e l s h w k n v g

1021 aatttttgge tataattcta ccgcggtttc acaagaatgy tcaggatcag ctgttttgaa 7ftf >
318 p i f g y n s t a v s q e w s g s a v l 6ftfi <

1081 cagtataac tctatccaat tattttatac aagggtagac acgtctgata acaataccaa
338 n s d n s i q l f y t r v d t s d n n t

1141 tcatcaaaaa attgctagcg ctactcttta tttactgat aataatggaa atgtatcact NheI
358 n h q k i a s a t l y l t d n n g n v s AC1(i)<>

1201 cgtcaggtga cgaaatgact atattgtatt tgaaggtgat ggctattact accaaactta AC2(i)<>
378 l a q v r n d y i v f e g d g y y y q t

1261 tgatcaatgg aaagctacta acaaaggtgc cgataatatt gcaatgcgtg atgctcatgt
398 y d q w k a t n k g a d n i a m r d a h

Fig 1 (3)

1321 aattgaagat ggtaatggg atcggtacct tgtttttgaa gcaagtactg gtttggaaaa
418 v i e d g n g d r y l v f e a s t g l e

1381 ttatcaaggc gaggaccaaa ttataactg gttaaattat ggcggagatg acgcatttaa
438 n y q g e d q i y n w l n y g g d d a f

1441 tatcaagagc ttatttagaa ttctttccaa tgatgatatt aagagtcggg caacttgggc
458 n i k s l f r i l s n d d i k s r a t w

1501 taatgcagct atcggtatcc tcaactaaa taaggacgaa aagaatccta aggtggcaga
478 a n a a i g i l k l n k d e k n p k v a

1561 gttatactca ccattaattt ctgcaccaat ggtaagcgat gaaattgagc gaccaaattg
498 e l y s p l i s a p m v s d e i e r p n

1621 agttaaatta ggtaataaat attacttatt tgccgtacc cgtttaaatc gaggaagtaa
518 v v k l g n k y y l f a a t r l n r g s

1681 tgatgatgct tggatgaatg ctaattatgc cyttggtgat aatgttgcaa tggtcggata
538 n d d a w m n a n y a v g d n v a m v g

1741 tgttgctgat agtctaactg gatcttataa gccattaaat gattctggag tagtcttgac
558 y v a d s l t g s y k p l n d s g v v l

1801 tgcttctggt cctgcaaact ggcggacagc aacttattca tattatgctg tccccgttgc
578 t a s v p a n w r t a t y s y y a v p v

1861 cggaaaagat gaccaagtat tagttacttc atatatgact aatagaaatg gagtagcggg
598 a g k d d q v l v t s y m t n r n g v a

1921 taaaggaatg gattcaactt gggcaecgag ttcttacta caaattaacc cggataaac 12ftfi <
618 g k g m d s t w a p s f i l q i n p d n

Fig 1 (4)

1981 aactactgtt ttagctaaaa tgactaatca aggggattgg atttgggatg attcaagcga

638 t t t v l a k m t n q g d w i w d d s s

2041 aaatcttgat atgattggtg atttagactc cgctgcttta cctggcgaac gtgataaacc

658 e n l d m i g d l d s a a l p g e r d k

2101 tgttgattgg gacttaattg gttatggatt aaaaccgcat gatcctgcta caccaaatga

678 p v d w d l i g y g l k p h d p a t p n

2161 tcctgaaacg ccaactacac cagaaacccc tgagacacct aatactccca aaacaccaaa

698 d p e t p t t p e t p e t p n t p k t p

2221 gactcctgaa aatcctggga caccctcaaac tcctaataca cctaatactc cggaaattcc

718 k t p e n p g t p q t p n t p n t p e i

2281 tttaactcca gaaacgccta agcaacctga aaccctaaact aataatcggt tgccacaaac

738 p l t p e t p k q p e t q t n n r l p q

2341 tggaaataat gccataaag ccattgattgg cctaggatatg ggaacattgc ttagtatggt

758 t g n n a n k a m i g l g m g t l l s m

2401 tggctcttgca gaaattaaca aacgtcgatt taactaaata ctttaaaata aaaccgctaa

778 f g l a e i n k r r f n

2461 gccttaaatt cagcttaacg gttttttatt ttaaaagttt ttattgtaaa aaagcgaatt

2521 atcattaata ctaatgcaat tgttgtaaga ccttacgaca gtagtaacaa tgaatttgcc

2581 catctttgtc gg

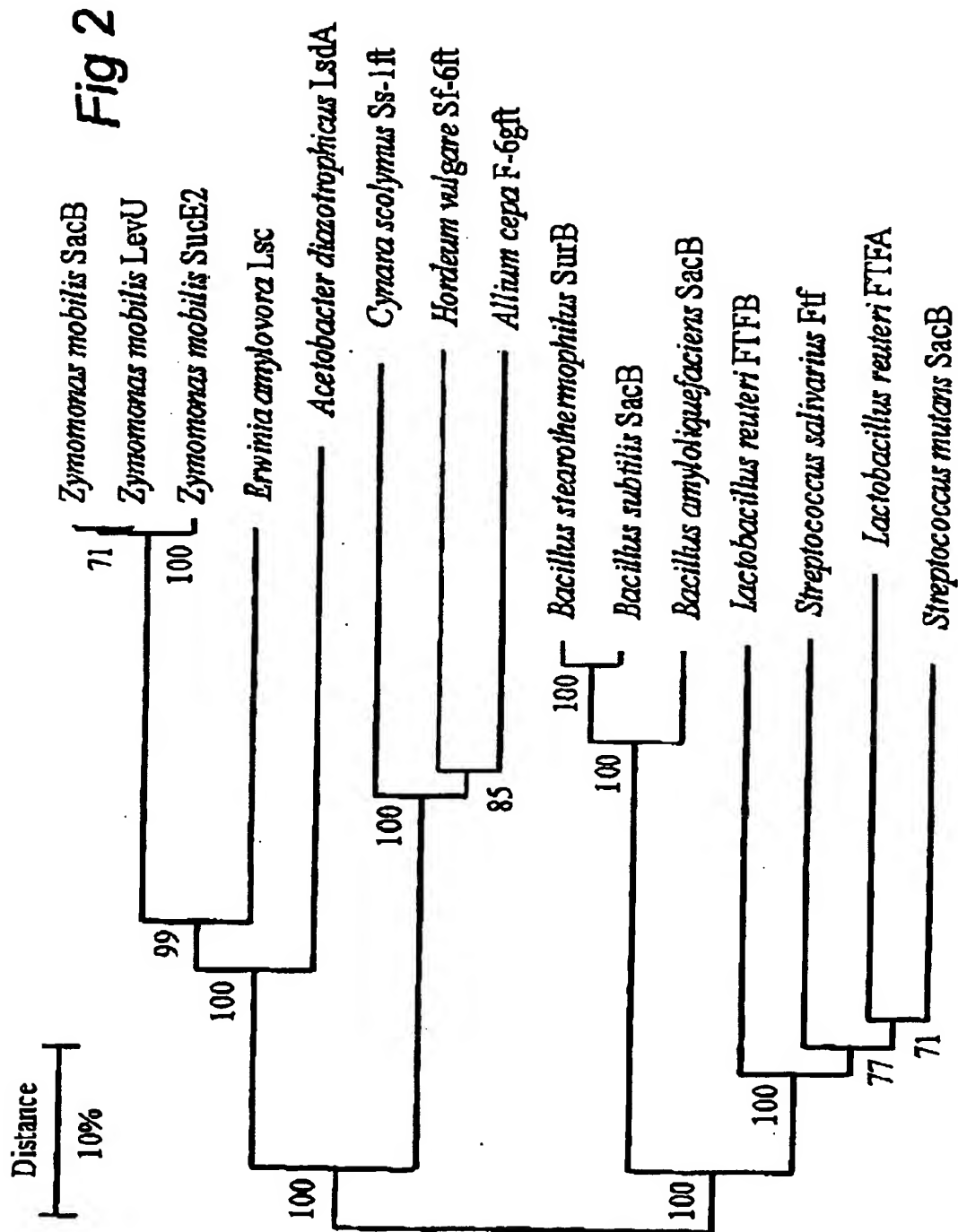


Fig 3

The N-terminal sequence of FTFB (levansucrase):

(A) Q V E S N N Y N G V A E V N T E R Q A N G Q I (G) (V) (D).

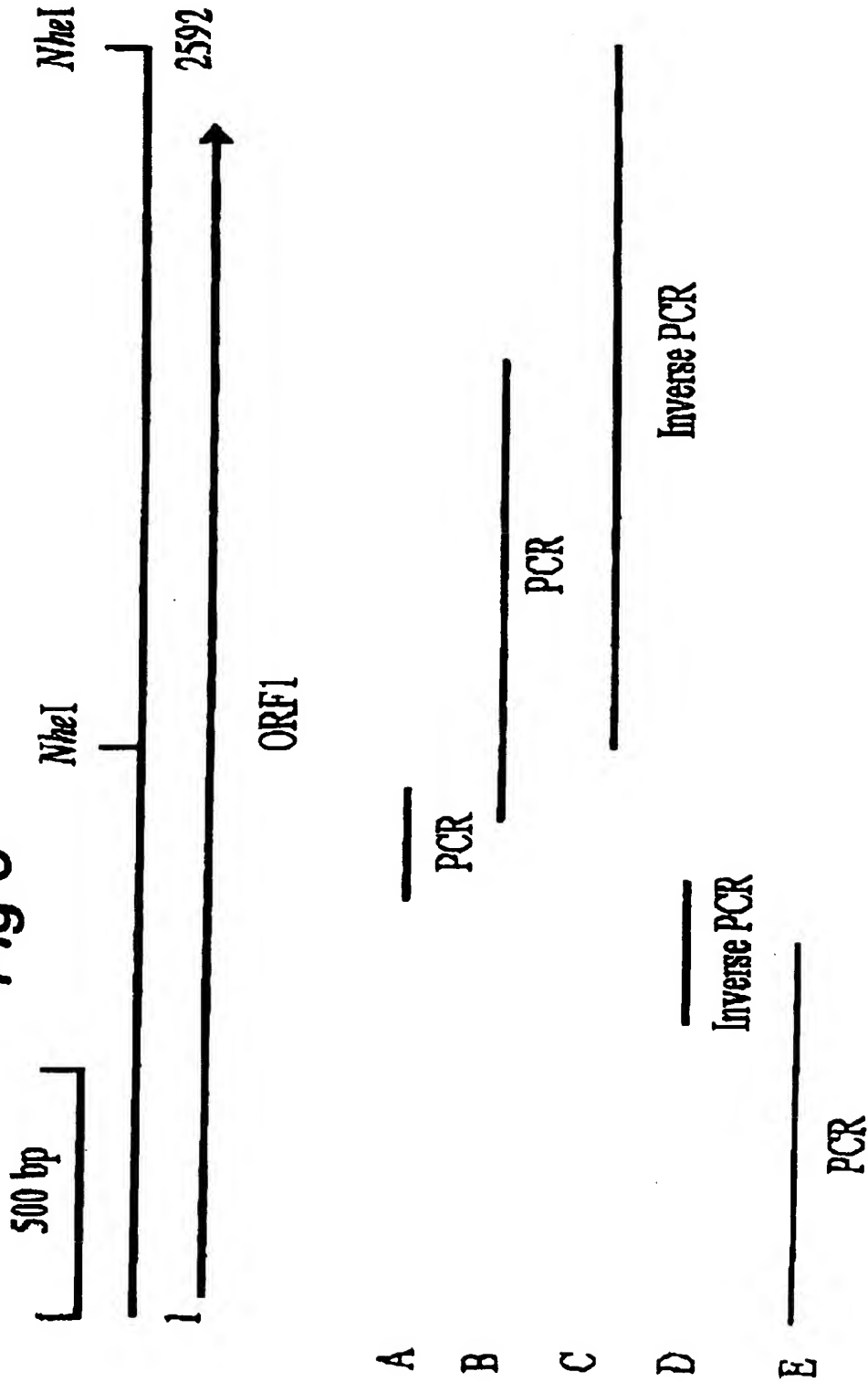
Internal peptide sequences of FTFB (levansucrase):

- (M) (A) H L D V W D S W P V Q D P (V),
- N A G S I F G T (K),
- V (E) (E) V Y S P K V S T L M A S D E V E.

Fig 4

5ftf			
<i>B. amyloliquefaciens</i> SacB	80	GLDVNDSEPLQAD	93
<i>B. subtilis</i> SacB	82	GLDVNDSEPLQAD	95
<i>S. mutans</i> SacB	243	DLDVNDSEPVQDAK	256
<i>S. salivarius</i> Ftf	282	EIDVNDSEPVQDAK	295
		:*****:*:*.	
6ftf			
<i>B. amyloliquefaciens</i> SacB	156	QTQENSGSATFTSDGK	171
<i>B. subtilis</i> SacB	158	QTQENSGSATFTSDGK	173
<i>S. mutans</i> SacB	312	LTQENSGSATVNEDEGS	327
<i>S. salivarius</i> Ftf	351	DDQQWSGSATVNSDGS	366
		*:*****:***.	
12ftf			
<i>B. amyloliquefaciens</i> SacB	440	KATFGPSFLMN	450
<i>B. subtilis</i> SacB	440	QSTFAPSFLN	450
<i>S. mutans</i> SacB	609	NSTWAPSFLIQ	619
<i>S. salivarius</i> Ftf	655	KSTWAPSFLIK	665
		:*:.*****:.	

Fig 5



FRUCTOSYLTRANSFERASES

CROSS-REFERENCE TO RELATED APPLICATION

This application is a continuation-in-part of application of U.S. application Ser. No. 09/604,958 filed on Jun. 28, 2000, now U.S. Pat. No. 6,635,460, which claims priority from European Application No. 00201872.9 filed on May 25, 2000.

The present invention is in the field of enzymatic production of biomolecules. The invention is particularly concerned with two novel fructosyltransferases derived from lactobacilli and with a process for recombinant production of the enzymes and for the production of useful levans, inulins and fructo-oligosaccharides from sucrose.

BACKGROUND OF THE INVENTION

Lactic acid bacteria (LAB) play an important role in the fermentative production of food and feed. Traditionally, these bacteria have been used for the production of for instance wine, beer, bread, cheese and yoghurt, and for the preservation of food and feed, e.g. olives, pickles, sausages, sauerkraut and silage. Because of these traditional applications, lactic acid bacteria are food-grade microorganisms that possess the Generally Recognised As Safe (GRAS) status. Due to the different products which are formed during fermentation with lactic acid bacteria, these bacteria contribute positively to the taste, smell and preservation of the final product. The group of lactic acid bacteria encloses several genera such as *Lactobacillus*, *Leuconostoc*, *Pediococcus*, *Streptococcus*, etc.

In recent years also the health promoting properties of lactic acid bacteria have received much attention. They produce an abundant variety of exopolysaccharides (EPS's). These polysaccharides are thought to contribute to human health by acting as prebiotic substrates, nutraceuticals, cholesterol lowering agents or immunomodulants.

To date high molecular weight polysaccharides produced by plants (such as cellulose, starch and pectin), seaweeds (such as alginate and carrageenan) and bacteria (such as alginate, gellan and xanthan) are used in several industrial applications as viscosifying, stabilising, emulsifying, gelling or water binding agents. Although all these polysaccharides are used as food additives, they originate from organisms not having the GRAS status. Thus they are less desirable than the exopolysaccharides of microorganisms, such as lactic acid bacteria, which have the GRAS status.

The exopolysaccharides produced by LAB can be divided in two groups, heteropolysaccharides and homopolysaccharides; these are synthesized by totally different mechanisms. The former consist of repeating units in which residues of different types of sugars are present and the latter consist of one type of monosaccharide. The synthesis of heteropolysaccharides by lactic acid bacteria, including lactobacilli, has been studied extensively in recent years. Considerably less information is available on the synthesis of homopolysaccharides from lactobacilli, although some studies have been performed. Homopolysaccharides with fructose as the constituent sugar can be divided into two groups, inulins and levans. Inulins consist of 2,1-linked β -fructofuranoside residues, whereas levans consist of 2,6-linked β -fructofuranoside residues. Both can be linear or branched. The size of bacterial levans can vary from 20 kDa up to several MDa. There is limited information on the synthesis of levans. In most detail this synthesis has been studied in *Zymomonas mobilis* and in *Bacillus* species.

Within lactic acid bacteria, fructosyltransferases have only been studied in streptococci. So far no fructosyltransferases have been reported in lactobacilli.

In a recent report the *Lactobacillus reuteri* strain LB 121 was found to produce both a glucan and a fructan when grown on sucrose, but only a fructan when grown on raffinose (van Geel-Schutten, G. H. et al., Appl. Microbiol. Biotechnol. (1998) 50, 697-703). In another report the glucan and fructan were characterised by their molecular weights (of 3,500 and 150 kDa respectively) and the glucan was reported to be highly branched with a unique structure consisting of a terminal, 4-substituted, 6-substituted, and 4,6-disubstituted α -glucose in a molar ratio 1.1: 2.7:1.5:1.0 (van Geel-Schutten, G. H. et al., Appl. Environ. Microbiol. (1999) 65, 3008-3014). The fructan was identified as a linear (2 \rightarrow 6)- β -D-fructofuranan (also called a levan). This was the first example of fructan synthesis by a *Lactobacillus* species.

SUMMARY OF THE INVENTION

Two novel genes encoding enzymes having fructosyltransferase activity have now been found in *Lactobacillus reuteri*, and their amino acid sequences have been determined. These are the first two enzymes identified in a *Lactobacillus* species capable of producing a fructan. One of the enzymes is an inulosucrase which produces a high molecular weight (>10⁷ Da) fructan containing β (2-1) linked fructosyl units and fructo-oligosaccharides, while the other is a levansucrase which produces a fructan containing β (2-6) linked fructosyl units. The invention thus pertains to the enzymes, to DNA encoding them, to recombinant cells containing such DNA and to their use in producing carbohydrates, as defined in the appending claims.

DESCRIPTION OF THE INVENTION

It was found according to the invention that one of the novel fructosyltransferases (FTFA; an inulosucrase) produces a high molecular weight inulin with β (2-1) linked fructosyl units and fructo-oligosaccharides. The fructo-oligosaccharides synthesis was also observed in certain *Lactobacillus* strains, in particular in certain strains of *Lactobacillus reuteri*. However, the inulin has not been found in *Lactobacillus reuteri* culture supernatants, but only in extracts of *E. coli* cells expressing the above-mentioned fructosyltransferase. This inulosucrase consists of either 798 amino acids (2394 nucleotides) or 789 amino acids (2367 nucleotides) depending on the potential start codon used. The molecular weight (MW) deduced of the amino acid sequence of the latter form is 86 kDa and its isoelectric point is 4.51, at pH 7.

The amino acid sequence of the inulosucrase is shown in SEQ ID No. 1 (FIG. 1, amino acid residues 1-789). As mentioned above, the nucleotide sequence contains two putative start codons leading to either a 2394 (see SEQ ID No. 3) or 2367 (see SEQ ID No. 2) nucleotide form of the inulosucrase. Both putative start codons are preceded by a putative ribosome binding site, GGGG (located 12 base pairs upstream its start codon) or AGGA (located 14 base pairs upstream its start codon), respectively (see FIG. 1 and SEQ ID No. 4).

The present invention covers a protein having inulosucrase activity with an amino acid identity of at least 65%, preferably at least 75%, and more preferably at least 85%, compared to the amino acid sequence of SEQ ID No. 1. The invention also covers a part of a protein with at least 15 contiguous amino acids which are identical to the corresponding part of the amino acid sequence of SEQ ID No. 1.

Fructosyltransferases have been found in several bacteria such as *Zymomonas mobilis*, *Erwinia amylovora*, *Acetobacter amylovora*, *Bacillus polymyxa*, *Bacillus amyloliquefaciens*, *Bacillus stearothermophilus*, and *Bacillus subtilis*. In lactic acid bacteria this type of enzyme previously has only been found in some streptococci. Most bacterial fructosyltransferases have a molecular mass of 50–100 kDa (with the exception of the fructosyltransferase found in *Streptococcus salivarius* which has a molecular mass of 140 kDa). Amino acid sequence alignment revealed that the novel inulosucrase of lactobacilli has high homology with fructosyltransferases originating from Gram positive bacteria, in particular with Streptococcus enzymes. The highest homology (FIG. 2) was found with the SacB enzyme of *Streptococcus mutans* Ingbritt A (62% identity within 539 amino acids).

Certain putative functions based on the alignment and site-directed mutagenesis studies can be ascribed to several amino acids of the novel inulosucrase. Asp-263, Glu-330, Asp-415, Glu-431, Asp-511, Glu-514, Arg-532 and/or Asp-551 of the amino acid sequence of SEQ ID No. 1 are identified as putative catalytic residues. Noteworthy, a hydrophobicity plot according to Kyte and Doolittle (1982) J. Mol. Biol. 157, 105–132 suggests that the novel inulosucrase contains a putative signal sequence according to the Von Heijne rule. The putative signal peptidase site is located between Gly at position 21 and Ala at position 22. Furthermore, it is striking that the C-terminal amino acid sequence of the novel inulosucrase contains a putative cell wall anchor amino acid signal LPXTG (SEQ ID No. 5) and a 20-fold repeat of the motif PXX (residues 690–749 of SEQ ID No. 1) (see figure 1.), where P is proline and X is any other amino acid. In 15 out of 20 repeats, however, the motif is PXT. This motif has so far not been reported in proteins of prokaryotic and eukaryotic origin.

A nucleotide sequence encoding any of the above mentioned proteins, mutants, variants or parts thereof is also a subject of the invention. Furthermore, the nucleic acid sequences corresponding to expression-regulating regions (promoters, enhancers, terminators) of at least 30 contiguous nucleic acids contained in the nucleic acid sequence (-67)-(-1) or 2367–2525 of SEQ ID No. 4 (see also FIG. 1) can be used for homologous or heterologous expression of genes. Such expression-regulating sequences are operationally linked to a polypeptide-encoding nucleic acid sequence such as the genes of the fructosyltransferase according to the invention. A nucleic acid construct comprising the nucleotide sequence operationally linked to an expression-regulating nucleic acid sequence is also covered by the invention.

A recombinant host cell, such as a mammalian (with the exception of human), plant, animal, fungal or bacterial cell, containing one or more copies of the nucleic acid construct mentioned above is an additional subject of the invention. The inulosucrase gene (starting at nucleotide 41) has been cloned in an *E. coli* expression vector under the control of an ara promoter in *E. coli* Top10. *E. coli* Top10 cells expressing the recombinant inulosucrase hydrolysed sucrose and synthesized fructan material. SDS-PAGE of arabinose induced *E. coli* Top10 cell extracts suggested that the recombinant inulosucrase has a molecular weight of 80–100 kDa, which is in the range of other known fructosyltransferases and in line with the molecular weight of 86 kDa deduced of the amino acid sequence depicted in FIG. 1.

The invention further covers an inulosucrase according to the invention which, in the presence of sucrose, produces a inulin having $\beta(2-1)$ -linked D-fructosyl units and fructo-

oligosaccharides. Two different types of fructans, inulins and levans, exist in nature. Surprisingly, the novel inulosucrase expressed in *E. coli* Top10 cell synthesizes a high molecular weight ($>10^7$ Da) inulin and fructo-oligosaccharides, while in *Lactobacillus reuteri* culture supernatants, in addition to the fructo-oligosaccharides, a levan and not an inulin is found. This discrepancy can have several explanations: the inulosucrase gene may be silent in *Lactobacillus reuteri*, or may not be expressed in *Lactobacillus reuteri* under the conditions tested, or the inulosucrase may only synthesize fructo-oligosaccharides in its natural host, or the inulin polymer may be degraded shortly after synthesis, or may not be secreted and remains cell-associated, or the inulosucrase may have different activities in *Lactobacillus reuteri* and *E. coli* Top10 cells.

It was furthermore found according to the invention that certain lactobacilli, in particular *Lactobacillus reuteri*, possess another fructosyltransferase, a levansucrase (FTFB), in addition to the inulosucrase described above. The N-terminal amino acid sequence of the fructosyltransferase purified from *Lactobacillus reuteri* supernatant was found to be QVESNNYNGVAEVNTERQANGQI (residues 2–24 of SEQ ID No. 6). Furthermore, three internal sequences were identified, namely (M)(A)HLDVWDSWPVQDP(V) (SEQ ID No. 7), NAGSIFGT(K) (SEQ ID No. 8), V(E) (E) VYSPKVSTLMASDEVE (SEQ ID No. 9). The N-terminal amino acid sequence could not be identified in the deduced inulosucrase sequence. Also the amino acid sequences of the three internal peptide fragments of the purified fructosyltransferase were not present in the putative inulosucrase sequence. Evidently, the inulosucrase gene does not encode the purified fructosyltransferase synthesizing the levan. The complete amino acid sequence of the levansucrase is shown in SEQ ID No. 11 and the nucleotide sequence is shown in SEQ ID No. 10. The levansucrase comprises a putative membrane anchor (see amino acids 761–765 in SEQ ID No. 11) and a putative membrane spanning domain (see amino acids 766–787 in SEQ ID No. 11). The fructan produced by the levansucrase was identified in the *Lactobacillus reuteri* culture supernatant as a linear (2 \rightarrow 6)- β -D-fructofuranan with a molecular weight of 150 kDa. The purified enzyme also produces this fructan.

Additionally, the invention thus covers a protein having levansucrase activity with an amino acid identity of at least 65%, preferably at least 75%, and more preferably at least 85%, compared to the amino acid sequence of SEQ ID No. 11. The second novel fructosyltransferase produces a high molecular weight fructan with $\beta(2-6)$ -linked fructosyl units with sucrose or raffinose as substrate. The invention also covers a part of a protein with least 15 contiguous amino acids, which are identical to the corresponding part of the amino acid sequence of SEQ ID No. 11. A nucleotide sequence encoding any of the above-mentioned proteins, mutants, variants or parts thereof is a subject of the invention as well as a nucleic acid construct comprising the nucleotide sequence mentioned above operationally linked to an expression-regulating nucleic acid sequence. A recombinant host cell, such as a mammalian (with the exception of human), plant, animal, fungal or bacterial cell, containing one or more copies of the nucleic acid construct mentioned above is an additional subject of the invention. The invention further covers a protein according to the invention which, in the presence of sucrose, produces a fructan having $\beta(2-6)$ -linked D-fructosyl units.

The invention also pertains to a process of producing an inulin-type and/or a levan-type of fructan as described above using fructosyltransferases according to the invention and a

suitable fructose source such as sucrose, stachyose or raffinose. The fructans may either be produced by *Lactobacillus* strains or recombinant host cells according to the invention containing one or both fructosyl transferases or by a fructosyltransferase enzyme isolated by conventional means from the culture of fructosyltransferase-positive *Lactobacilli*, especially a *Lactobacillus reuteri*, or from a recombinant organism containing the fructosyltransferase gene or genes.

Additionally, the invention concerns a process of producing fructo-oligosaccharides containing the characteristic structure of the fructans described above using a *Lactobacillus* strain or a recombinant host cell according to the invention containing one or both fructosyltransferases or an isolated fructosyltransferase according to the invention. There is a growing interest in oligosaccharides derived from homopolysaccharides, for instance for prebiotic purposes. Several fructo- and gluco-oligosaccharides are known to stimulate the growth of bifidobacteria in the human colon. Fructo-oligosaccharides produced by the fructosyltransferase described above are also part of the invention. Another way of producing fructo-oligosaccharides is by hydrolysis of the fructans described above. This hydrolysis can be performed by known hydrolysis methods such as enzymatic hydrolysis with enzymes such as levanase or inulinase or by acid hydrolysis. The fructo-oligosaccharides can also be produced in the presence of a fructosyltransferase according to the invention and an acceptor molecule such as lactose or maltose. The fructo-oligosaccharides to be produced according to the invention preferably contain at least 2, more preferably at least 3, up to about 20 anhydrofructose units, optionally in addition to one or more other (glucose, galactose, etc.) units. These fructo-oligosaccharides are useful as prebiotics, and can be administered to a mammal in need of improving the bacterial status of the colon.

The invention also concerns chemically modified fructans and fructo-oligosaccharides based on the fructans described above. Chemical modification can be achieved by oxidation, such as hypochlorite oxidation resulting in ring-opened 2,3-dicarboxy-anhydrofructose units (see e.g. EP-A-427349), periodate oxidation resulting in ring-opened 3,4-dialdehyde-anhydrofructose units (see e.g. WO 95/12619), which can be further oxidised to (partly) carboxylated units (see e.g. WO 00/26257), TEMPO-mediated oxidation resulting in 1- or 6-carboxy-anhydrofructose units (see e.g. WO 95/07303). The oxidised fructans have improved water-solubility, altered viscosity and a retarded fermentability and can be used as metal-complexing agents, detergent additives, strengthening additives, bioactive carbohydrates, emulsifiers and water binding agents. They can also be used as starting materials for further derivatisation such as cross-linking and the introduction of hydrophobes. Oxidised fructans coupled to amino compounds such as proteins, or fatty acids can be used as emulsifiers and stabilizers. (Partial) hydrolysis of fructans according to the invention and modified fructans according to the invention results in fructo-oligosaccharides, which can be used as bioactive carbohydrates or prebiotics. The oxidised fructans of the invention preferably contain 0.05–1.0 carboxyl groups per anhydrofructose unit, e.g. as 6- or 1-carboxyl units.

Another type of chemical modification is phosphorylation, as described in O.B. Wurzburg (1986) Modified Starches: properties and uses. CRC Press Inc., Boca Raton, 97–112. One way to achieve this modification is by dry heating fructans with a mixture of monosodium and disodium hydrogen phosphate or with tripolyphosphate. The phosphorylated fructans are suitable as wet-end additives in

papermaking, as binders in paper coating compositions, as warp sizing-agents, and as core binders for sand molds for metal casting. A further type of derivatisation of the fructans is acylation, especially acetylation using acetic or propionic anhydride, resulting in products suitable as bleaching assistants and for the use in foils. Acylation with e.g. alkenyl succinic anhydrides or (activated) fatty acids results in surface-active products suitable as e.g. surfactants, emulsifiers, and stabilizers.

Hydroxyalkylation, carboxymethylation, and aminoalkylation are other methods of chemical derivatisation of the fructans. Hydroxyalkylation is commonly performed by base-catalysed reaction with alkylene oxides, such as ethylene oxide, propylene oxide or epichlorohydrine; the hydroxyalkylated products have improved solubility and viscosity characteristics. Carboxymethylation is achieved by reaction of the fructans with mono-chloroacetic acid or its alkali metal salts and results in anionic polymers suitable for various purposes including crystallisation inhibitors, and metal complexants. Amino-alkylation can be achieved by reaction of the fructans with alkylene imines, haloalkyl amines or amino-alkylene oxides, or by reaction of epichlorohydrine adducts of the fructans with suitable amines. These products can be used as cationic polymers in a variety of applications, especially as a wet-end additive in paper making to increase strength, for filler and fines retention, and to improve the drainage rate of paper pulp. Other potential applications include textile sizing and wastewater purification. The above mentioned modifications can be used either separately or in combination depending on the desired product. Furthermore, the degree of chemical modification is variable and depends on the intended use. If necessary 100% modification, i.e. modification of all anhydrofructose units can be performed. However, partial modification, e.g. from 1 modified anhydrofructose unit per 100 up to higher levels, will often be sufficient in order to obtain the desired effect. The modified fructans have a DP (degree of polymerisation) of at least 100, preferably at least 1000 units.

Use of a *Lactobacillus* strain capable of producing a levan, inulin or fructo-oligosaccharides or a mixture thereof, as a probiotic, is also covered by the invention. Preferably, the *Lactobacillus* strain is also capable of producing a glucan, especially an 1,4/1,6- α -glucan as referred to above. The efficacy of some *Lactobacillus reuteri* strains as a probiotics has been demonstrated in various animals such as for instance poultry and humans. The administration of some *Lactobacillus reuteri* strains to pigs resulted in significantly lower serum total and LDL-cholesterol levels, while in children *Lactobacillus reuteri* is used as a therapeutic agent against acute diarrhea. For this and other reasons *Lactobacillus reuteri* strains, which were not reported to produce the glucans or fructans described herein, have been supplemented to commercially available probiotic products. The mode of action of *Lactobacillus reuteri* as a probiotic is still unclear. Preliminary studies indicated that gut colonization by *Lactobacillus reuteri* may be of importance. According to the invention, it was found that the mode of action of *Lactobacillus reuteri* as a probiotic may reside partly in the ability to produce polysaccharides. *Lactobacillus* strains, preferably *Lactobacillus reuteri* strains, and more preferably *Lactobacillus reuteri* strain LB 121 and other strains containing one or more fructosyltransferase genes encoding proteins capable of producing inulins, levans and/or fructo-oligosaccharides can thus advantageously be used as a probiotic. They can also, together with these polysaccharides, be used as a symbiotic (instead of the term symbiotic, the term synbiotic can also be used). In that

respect another part of the invention concerns a probiotic or symbiotic composition containing a *Lactobacillus* strain capable of producing an inulin, a levan or fructo-oligosaccharides and/or a glucan or a mixture thereof, said production being performed according to the process according to the invention. The probiotic or symbiotic compositions of the invention may be directly ingested with or without a suitable vehicle or used as an additive in conjunction with foods. They can be incorporated into a variety of foods and beverages including, but not limited to, yoghurts, ice creams, cheeses, baked products such as bread, biscuits and cakes, dairy and dairy substitute foods, confectionery products, edible oil compositions, spreads, breakfast cereals, juices and the like.

Furthermore, the invention pertains to a process of improving the microbial status in the mammalian colon comprising administering an effective amount of a *Lactobacillus* strain capable of producing an oligosaccharide or polysaccharide according to the invention and to a process of improving the microbial status of the mammalian colon comprising administering an effective amount of an oligosaccharide or polysaccharide produced according to the process according to the invention.

EXAMPLES

EXAMPLE 1

Isolation of DNA from *Lactobacillus reuteri* Nucleotide Sequence Analysis of the Inulosucrase (fitA) Gene, Construction of Plasmids for Expression of the Inulosucrase Gene in *E. coli* Top10 Expression of the Inulosucrase gene in *E. coli* Top10 and Identification of the Produced Polysaccharides Produced by the Recombinant Enzyme.

General procedures for cloning, DNA manipulations and agarose gel electrophoresis were essentially as described by Sambrook et al. (1989) Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, N.Y. Restriction endonuclease digestions and ligations with T4 DNA ligase were performed as recommended by the suppliers. DNA was amplified by PCR techniques using ampliTaq DNA polymerase (Perkin Elmer) or Pwo DNA polymerase. DNA fragments were isolated from agarose gels using the Qiagen extraction kit (Qiagen GmbH), following the instructions of the suppliers. *Lactobacillus reuteri* strain 121 (LMG 18388) was grown at 37° C. in MRS medium (DIFCO) or in MRS-s medium (MRS medium containing 100 g/l sucrose instead of 20 g/l glucose). When fructo-oligosaccharides production was investigated phosphate was omitted and ammonium citrate was replaced by ammonium nitrate in the MRS-s medium. *E. coli* strains were grown aerobically at 37° C. in LB medium, where appropriate supplemented with 50 µg/ml ampicillin (for selection of recombinant plasmids) or with 0.02% (w/v) arabinose (for induction of the inulosucrase gene).

Total DNA of *Lactobacillus reuteri* was isolated according to Verhasselt et al. (1989) FEMS Microbiol. Lett. 59, 135-140 as modified by Nagy et al. (1995) J. Bacteriol. 177, 676-687.

The inulosucrase gene was identified by amplification of chromosomal DNA of *Lactobacillus reuteri* with PCR using degenerated primers (5 fitf, 6 fiti, and 12 fiti, see table 1) based on conserved amino acid sequences deduced from different bacterial fructosyltransferase genes (SacB of *Bacillus amyloliquefaciens*, SacB of *Bacillus subtilis*, *Streptococcus mutans* fructosyltransferase and *Streptococcus salivarius* fructosyltransferase, see FIG. 4) and *Lactobacillus*

reuteri DNA as template. Using primers 5 fitf and 6 fiti, an amplification product with the predicted size of about 234 bp was obtained (FIG. 5A). This 234 bp fragment was cloned in *E. coli* JM109 using the pCR2.1 vector and sequenced. Transformations were performed by electroporation using the BioRad gene pulser apparatus at 2.5 kV, 25 µF and 200 Ω, following the instructions of the manufacturer. Sequencing was performed according to the method of Sanger et al. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467. Analysis of the obtained sequence data confirmed that part of a fructosyltransferase (fitf gene had been isolated. The 234 bp amplified fragment was used to design primers 7 fitf and 8 fiti (see table 1). PCR with the primers 7 fitf and 12 fiti gave a product of the predicted size of 948 bp (see FIG. 5B); its sequence showed clear similarity with previously characterized fructosyltransferase genes. The 948 bp amplified fragment was used to design the primers fitAC1(i) and fitAC2(i) (see table 1) for inverse PCR. Using inverse PCR techniques a 1438 bp fragment of the inulosucrase gene was generated, including the 3' end of the inulosucrase gene (see FIG. 5C). The remaining 5' fragment of the inulosucrase gene was isolated with a combination of standard and inverse PCR techniques. Briefly, *Lactobacillus reuteri* DNA was cut with restriction enzyme XhoI and ligated. PCR with the primers 7 fitf and 8 fiti, using the ligation product as a template, yielded a 290 bp PCR product which was cloned into pCR2.1 and sequenced. This revealed that primer 8 fiti had annealed aspecifically as well as specifically yielding the 290 bp product (see FIG. 5D).

At this time, the N-terminal amino acid sequence of a fructosyltransferase enzyme (FTFB) purified from the *Lactobacillus reuteri* strain 121 was obtained. This sequence consisted of the following 23 amino acids: QVESNNYN-GVAEVNTERQANGQI (residues 2-24 of SEQ ID No. 6). The degenerated primer 19 fitf (YNGVAEV) (residues 8-14 of SEQ ID NO: 6) was designed on the basis of a part of this N-terminal peptide sequence and primer 20 fiti was designed on the 290 bp PCR product. PCR with primers 19 fitf and 20 fiti gave a 754 bp PCR product (see FIG. 5E), which was cloned into pCR2.1 and sequenced. Both DNA strands of the entire fructosyltransferase gene were double sequenced. In this way the sequence of a 2.6 kb region of the *Lactobacillus reuteri* DNA, containing the inulosucrase gene and its surroundings were obtained.

The plasmids for expression of the inulosucrase gene in *E. coli* Top10 were constructed as described hereafter. A 2414 bp fragment, containing the inulosucrase gene starting at the first putative start codon at position 41, was generated by PCR, using primers fitA1 and fitA2i. Both primers contained suitable restriction enzyme recognition sites (a NcoI site at the 5' end of fitA1 and a BglII site at the 3' end of fitA2i). PCR with *Lactobacillus reuteri* DNA, Pwo DNA polymerase and primers fitA1 and fitA2i yielded the complete inulosucrase gene flanked by NcoI and BglII restriction sites. The PCR product with blunt ends was ligated directly into pCRbluntII-Topo. Using the NcoI and BglII restriction sites, the putative fitA gene was cloned into the expression vector pBAD, downstream of the inducible arabinose promoter and in frame upstream of the Myc epitope and the His tag. The pBAD vector containing the inulosucrase gene (pSVH101) was transformed to *E. coli* Top10 and used to study inulosucrase expression. Correct construction of plasmid containing the complete inulosucrase gene was confirmed by restriction enzyme digestion analysis and by sequence analysis, showing an in frame cloning of the inulosucrase gene using the ribosomal binding site provided by the pBAD vector and the first putative start codon (at position 41) of inulosucrase (see FIG. 1).

Plasmid DNA of *E. coli* was isolated using the alkaline lysis method of Birnboim and Doly (1979) Nucleic Acids Res. 7, 1513–1523 or with a Qiagen plasmid kit following the instructions of the supplier. Cells of *E. coli* Top10 with pSVH101 were grown overnight in LB medium containing 0.02% (w/v) arabinose and were harvested by centrifugation. The pellet was washed with 25 mM sodium acetate buffer pH 5.4 and the suspension was centrifuged again. Pelleted cells were resuspended in 25 mM sodium acetate buffer pH 5.4. Cells were broken by sonication. Cell debris and intact cells were removed by centrifugation for 30 min at 4° C. at 10,000×g and the resulting cell free extract was used in the enzyme assays.

The fructosyltransferase activities were determined at 37° C. in reaction buffer (25 mM sodium acetate, pH 5.4, 1 mM CaCl₂, 100 g/l sucrose) by monitoring the release of glucose from sucrose, by detecting fructo-oligosaccharides or by determining the amount of fructan polymer produced using *E. coli* cell free extracts or *Lactobacillus reuteri* culture supernatant as enzyme source. Sucrose, glucose and fructose were determined enzymatically using commercially available kits.

Fructan production by *Lactobacillus reuteri* was studied with cells grown in MRS-s medium. Product formation was also studied with cell-free extracts of *E. coli* containing the novel inulosucrase incubated in reaction buffer (1 mg protein/10 ml buffer, incubated overnight at 37° C.). Fructans were collected by precipitation with ethanol. ¹H-NMR spectroscopy and methylation analysis were performed as described by van Geel-Schutten et al. (1999) Appl. Environ. Microbiol. 65, 3008–3014. The molecular weights of the fructans were determined by high performance size exclusion chromatography coupled on-line with a multi angle laser light scattering and a differential refractive index detector. Fructo-oligosaccharide synthesis was studied in *Lactobacillus reuteri* culture supernatants and in extracts of *E. coli* cells containing the novel inulosucrase incubated in reaction buffer (1 mg protein/10 ml buffer, incubated overnight at 37° C.). Glucose and fructose were determined enzymatically as described above and fructo-oligosaccharides produced were analyzed using a Dionex column. The incubation mixtures were centrifuged for 30 min at 10,000×g and diluted 1:5 in a 100% DMSO solution prior to injection on a Dionex column. A digest of inulin (DPI-20) was used as a standard. Separation of compounds was achieved with anion-exchange chromatography on a CarboPac Pa1 column (Dionex) coupled to a CarboPac PA1 guard column (Dionex). Using a Dionex GP50 pump the following gradient was generated: % eluent B is 5% (0 min); 35% (10 min); 45% (20 min); 65% (50 min); 100% (54–60 min); 5% (61–65 min). Eluent A was 0.1 M NaOH and eluent B was 0.6 M NaAc in a 0.1 M NaOH solution. Compounds were detected using a Dionex ED40 electrochemical detector with an AU working electrode and a Ag/AgCl reference-electrode with a sensitivity of 300 nC. The pulse program used was: +0.1 Volt (0–0.4 s); +0.7 Volt (0.41–0.60 s); –0.1 Volt (0.61–1.00 s). Data were integrated using a Perkin Elmer Turbochrom data integration system. A different separation of compounds was done on a cation exchange column in the calcium form (Benson BCX4). As mobile phase Ca-EDTA in water (100 ppm) was used. The elution speed was 0.4 ml/min at a column temperature of 85° C. Detection of compounds was done by a refractive index (Jasco 830-RI) at 40° C. Quantification of compounds was achieved by using the software program Turbochrom (Perkin Elmer).

SDS-PAGE was performed according to Laemmli (1970) Nature 227, 680–685 using 7.5% polyacrylamide gels. After

electrophoresis gels were stained with Coomassie Brilliant Blue or an activity staining (Periodic Acid Schiff, PAS) was carried out as described by Van Geel-Schutten et al. (1999) Appl. Environ. Microbiol. 65, 3008–3014.

TABLE 1

Nucleotide sequence of primers used in PCR reactions to identify the inulosucrase gene.

Primer name	Location (bp)	Nucleotide sequence (and SEQ ID No)
ftfAC1	1176	CTG-ATA-ATA-ATG-GAA-ATG-TAT-CAC (SEQ ID No. 12)
ftfAC2i	1243	CAT-GAT-CAT-AAG-TTT-GGT-AGT-AAT-AG (SEQ ID No. 13)
ftfAC1	1176	GTG-ATA-CAT-TTC-CAT-TAT-TAT-CAG (SEQ ID No. 14)
ftfAC2	1243	CTA-TTA-CTA-CCA-AAC-TTA-TGA-TCA-TG (SEQ ID No. 15)
ftfA1		CCA-TGG-CCA-TGG-TAG-AAC-GCA-AGG-AAC-ATA-AAA-AAA-TG (SEQ ID No. 16)
ftfA2i		AGA-TCT-AGA-TCT-GTT-AAA-TCG-ACG-TTT-GTT-AAT-TTC-TG (SEQ ID No. 17)
5ftf	845	GAY-GTN-TGG-GAY-WSN-TGG-GCC (SEQ ID No. 18)
6ftf	1052	GTN-GCN-SWN-CCN-SWC-CAY-TSY-TG (SEQ ID No. 19)
7ftf	1009	GAA-TGT-AGG-TCC-AAT-TTT-TGG-C (SEQ ID No. 20)
8ftf	864	CCT-GTC-CGA-ACA-TCT-TGA-ACT-G (SEQ ID No. 21)
12ftf	1934	ARR-AAN-SWN-GGN-GCV-MAN-GTN-SW (SEQ ID No. 22)
19ftf	1	TAY-AAY-GGN-GTN-GCN-GAR-GTN-AA (SEQ ID No. 23)
20ftf	733	CCG-ACC-ATC-TTG-TTT-GAT-TAA-C (SEQ ID No. 24)

Listed from left to right are: primer name (i, inverse primer), location (in bp) in ftfA and the sequence from 5' to 3' according to IUB group codes (N = any base; M = A or C; R = A or G; W = A or T; S = C or G; Y = C or T; K = G or T; B = not A; D = not C; H = not G; and V = not T).

EXAMPLE 2

Purification and Amino Acid Sequencing of the Levansucrase (FTFB).

Protein Purification

Samples were taken between each step of the purification process to determine the enzyme activity (by glucose GOD-Perid method) and protein content (by Bradford analysis and acrylamide gel electrophoresis). Collected chromatography fractions were screened for glucose liberating activity (GOD-Perid method) to determine the enzyme activity.

One liter of an overnight culture of LB121 cells grown on MRS medium containing 50 grams per liter maltose was centrifuged for 15 min. at 10,000×g. The supernatant was precipitated with 1.5 liter of a saturated ammonium sulphate solution. The ammonium sulphate solution was added at a rate of 50 ml/min. under continuous stirring. The resulting 60% (w/v) ammonium sulphate solution was centrifuged for 15 min. at 10,000×g. The precipitate was resuspended in 10 ml of a sodium phosphate solution (10 mM, pH 6.0) and dialysed overnight against 10 mM sodium phosphate, pH 6.0.

A hydroxylapatite column was washed with a 10 mM sodium phosphate solution pH 6.0; the dialysed sample was loaded on the column. After eluting the column with 200 mM sodium phosphate, pH 6.0 the eluted fractions were screened for glucose releasing activity and fractions were pooled for phenyl superose (a hydrophobic interactions column) chromatography. The pooled fractions were diluted

1:1 (v:v) with 25 mM sodium acetate, 2 M ammonium sulphate, pH 5.4 and loaded on a phenyl superose column (washed with 25 mM sodium acetate, 1 M ammonium sulphate, pH 5.4). In a gradient from 25 mM sodium acetate, 1 M ammonium sulphate, pH 5.4 (A) to 25 mM sodium acetate, pH 5.4 (B) fractions were collected from 35% B to 50% B.

Pooled fractions from the phenyl superose column were loaded on a gel filtration (superdex) column and eluted by a 25 mM acetate, 0.1 M sodium chloride, pH 5.4 buffer. The superdex fractions were loaded on a washed (with 25 mM sodium acetate, pH 5.4) Mono Q column and eluted with 25 mM sodium acetate, 1 M sodium chloride, pH 5.4. The fractions containing glucose liberating activity were pooled, dialysed against 25 mM sodium acetate, pH 5.4, and stored at -20° C.

A levansucrase enzyme was purified from LB121 cultures grown on media containing maltose using ammonium sulphate precipitation and several chromatography column steps (table 2). Maltose (glucose—glucose) was chosen because both glucanucrase and levansucrase can not use maltose as substrate. LB121 will grow on media containing maltose but will not produce polysaccharide. From earlier experiments it was clear that even with harsh methods the levansucrase enzyme could not be separated from its product levan. These harsh methods included boiling the levan in a SDS solution and treating the levan with HCl and TFA. No levansucrase enzyme was commercially available for the enzymatic breakdown of levan. Only a single levansucrase was detected in maltose culture supernatants. In order to prove that the enzyme purified from maltose culture supernatant is the same enzyme which is responsible for the levan production during growth on raffinose, biochemical and biophysical tests were performed.

TABLE 2

Purification of the <i>Lactobacillus reuteri</i> LB 121 levansucrase (FTFB) enzyme					
Step	Protein (mg)	Total Activity (U)	Specific Activity (U/mg)	Purification (fold)	Yield (%)
Supernatant	128	64	0.5	1	100
Ammonium sulfate precipitation (65%)	35.2	42	1.2	2.4	65.6
Hydroxyl apatite	1.5	30.6	20.4	40.8	47.8
Phenyl superose	0.27	23	85	170	36
Gel	0.055	10	182	360	16
Filtration MonoQ	0.0255	4	176	352	6

Amino Acid Sequencing of FTFB

A 5% SDS-PAA gel was allowed to "age" overnight in order to reduce the amount of reacting chemical groups in the gel. Reaction of chemicals in the PAA gel (TEMED and ammonium persulphate) with proteins can cause some undesired effects, such as N-terminal blocking of the protein, making it more difficult to determine the protein amino acid composition. 0.1 mM thioglycolic acid (scavenger to reduce the amount of reactive groups in the PAA gel material) was added to the running buffer during electrophoresis.

In order to determine the amino acid sequence of internal peptides of protein bands running in a SDS-PAA gel, protein containing bands were cut out of the PAA gel. After frac-

tionating the protein by digestion with chymotrypsin the N-terminal amino acid sequences of the digested proteins were determined (below).

N-terminal sequencing was performed by Western blotting of the proteins from the PAA gel to an Immobilon PVDF membrane (Millipore/Waters Inc.) at 0.8 mA/cm² for 1 h. After staining the PVDF membrane with Coomassie Brilliant Blue without adding acetic acid (to reduce N-terminal blocking) and destaining with 50% methanol, the corresponding bands were cut out of the PVDF membrane for N-terminal amino acid sequence determination.

Amino acid sequence determination was performed by automated Edman degradation as described by Koningsberg and Steinman (1977) The proteins (third edition) volume 3, 1-178 (Neurath and Hill, eds.). The automated equipment for Edman degradation was an Applied Biosystems model 477A pulse-liquid sequenator described by Hewick et al. (1981), J. Biol. Chem. 15, 7990-7997 connected to a RP-HPLC unit (model 120A, Applied Biosystems) for amino acid identification.

The N-terminal sequence of the purified FTFB was determined and found to be: (A) Q V E S N N Y N G V A E V N T E R Q A N G Q I (G) (V) (D) (SEQ ID No. 6). Three internal peptide sequences of the purified FTFB were determined: (M) (A) H L D V W D S W P V Q D P (V) (SEQ ID No. 7); N A G S I F G T (K) (SEQ ID No. 8); and V (E) (E) V Y S P K V S T L M A S D E V E (SEQ ID No. 9).

The following primers were designed on the basis of the N-terminal and internal peptide fragments of FTFB. Listed from left to right are: primer name, source peptide fragment and sequence (from 5' to 3'). FTFB1+FTFB3i yields approximately a 1400 bp product in a PCR reaction. FTFB1 forward (N-terminal): AA T/C-TAT-AA T/C-GG T/C-GTT-GC G/A-T/C GA-AGT (SEQ ID No. 25); and FTFB3i reverse (Internal 3): TAC-CGN-A/T C/G N-CTA-CTT-CAA-C/T (SEQ ID No. 26). The FTFB gene was partly isolated by PCR with primers FTFB1 and FTFB3i. PCR with these primers yielded a 1385 bp amplicon, which after sequencing showed high homology to fitA and SacB from *Streptococcus mutans*.

EXAMPLE 3

Oxidation of Levans

For TEMPO-mediated oxidation, a levan according to the invention prepared as described above (dry weight 1 g, 6.15 mmol) was resuspended in 100 ml water. Next, 2,2,6,6-tetramethylpiperidine-1-oxyl (TEMPO; 1% by weight compared to the polysaccharide (0.01 g, 0.065 mmol)) was added and resuspended in 20 min. Sodium bromide (0.75 g, 7.3 mmol) was added and the suspension was cooled down to 0° C. This reaction also proceeded without bromide. A solution of hypochlorite (6 ml, 15% solution, 12.6 mmol) was adjusted to pH 10.0 with 3M HCl and cooled to 0° C. This solution was added to the suspension of the polysaccharide and TEMPO. The course of the reaction was followed by monitoring the consumption of sodium hydroxide solution, which is equivalent to the formation of uronic acid. After 30 min, 60 ml 0.1M NaOH was consumed. This amount corresponds to the formation of 97% uronic acid. Thereafter, the solution was poured out in 96% ethanol (comprising 70% of the volume of the solution) causing the product to precipitate. The white precipitate was centrifuged, resuspended in ethanol/water (70/30 v/v) and centrifuged again. Next, the precipitate was resuspended in 96% ethanol and centrifuged. The obtained product was dried at reduced pressure. The uronic acid content was determined by means of the uronic acid assay according to Blumenkrantz and Abdoe-Hansen (Anal. Biochem., 54

(1973), 484). A calibration curve was generated using polygalacturonic acid (5, 10, 15 and 20 μg). With this calibration curve the uronic acid content in a sample of 20 μg of the product was determined. The obtained result was a content of 95% uronic acid with a yield of 96%.

Partial Oxidation

For partial oxidation, a levan according to the invention (dry weight 2 g, 12.3 mmol) was resuspended in 25 ml water. Next, TEMPO (1% by weight compared to the polysaccharide (0.02 g, 0.13 mmol)) was added, resuspended in 20 min and cooled to 0° C. A solution of hypochlorite (1 ml, 15% solution, 2.1 mmol) was adjusted to pH 9.0 with 3M HCl and cooled down to 0° C. This solution was added to the suspension of the polysaccharide and TEMPO. Within 5 min the mixture became a solid gel.

EXAMPLE 4

Adhesion of *Lactobacillus reuteri* Strains to Caco-2 Cell Lines

The adhesion of *Lactobacillus reuteri* strains to Caco-2 cell lines was determined as described below. Firstly, a bacterial suspension was prepared as follows. *Lactobacillus reuteri* strains LB 121, 35-5, K24 and DSM20016 and *L. rhamnosus* LGG (a well known probiotic strain with good adhering properties) were cultured in MRS broth supplemented with 5 $\mu\text{l/ml}$ of methyl-1,2-[^3H]-thymidine at 37° C. for 18–20 h before the adhesion assays. The cultures were harvested by centrifugation, washed with phosphate buffered saline (PBS) and resuspended in PBS or PBS supplemented with 30 g/l sucrose (see Table 3) to a final density of about 2×10^9 cfu/ml. Prior to the adhesion assay, the cell suspensions in PBS with 30 g/l sucrose were incubated for 1 hour at 37° C., whereas the cell suspensions in PBS were kept on ice for 1 hour. After incubation at 37° C., the suspensions in PBS with sucrose were centrifuged and the cells were washed with and resuspended in PBS to a final density of about 2×10^9 cfu/ml.

Caco-2 cells were cultured as follows. Subcultures of Caco-2 cells (ATCC, code HTB 37, human colon adenocarcinoma), stored as frozen stock cultures in liquid nitrogen were used for the adhesion tests. The Caco-2 cells were grown in culture medium consisting of Dulbecco's modified Eagle medium (DMEM), supplemented with heat-inactivated foetal calf serum (10% v/v), non-essential amino acids (1% v/v), L-glutamine (2 mM) and gentamicin (50 $\mu\text{g/ml}$). About 2,000,000 cells were seeded in 75 cm^2 tissue culture flasks containing culture medium and cultured in a humidified incubator at 37° C. in air containing 5% CO_2 . Near confluent Caco-2 cell cultures were harvested by trypsinisation and resuspended in culture medium. The number of cells was established using a Bürker-Türk counting chamber.

TABLE 3

Incubation of the different <i>Lactobacillus</i> strains prior to the adhesion assays.			
Lactobacillus strain	Extra incubation	Polysaccharide produced	Group
reuteri 121	PBS sucrose, 37° C. for 1 hr	glucan and fructan	As
reuteri 35-5	PBS sucrose, 37° C. for 1 hr	glucan	Bs
reuteri K24	PBS sucrose, 37° C. for 1 hr	none	Cs

TABLE 3-continued

Incubation of the different <i>Lactobacillus</i> strains prior to the adhesion assays.			
Lactobacillus strain	Extra incubation	Polysaccharide produced	Group
reuteri 121	PBS on ice	none	D
reuteri	PBS on ice	none	E
DSM20016*	PBS on ice	none	F
rhamnosus GG	PBS on ice	none	F

*Type strain of *L. reuteri*

For the following experiments a Caco-2 monolayer transport system was used. Caco-2 cells cultured in a two-compartment transport system are commonly used to study the intestinal, epithelial permeability. In this system the Caco-2 cell differentiates into polarized columnar cells after reaching confluency. The Caco-2 system has been shown to simulate the passive and active transcellular transport of electrolytes, sugars, amino acids and lipophilic compounds (Hillgren et al. 1995, Dulfer et al., 1996, Duizer et al., 1997). Also, a clear correlation between the in vivo absorption and the permeability across the monolayers of Caco-2 cells has been reported (Artursson and Karlsson, 1990). For the present transport studies, Caco-2 cells were seeded on semi-permeable filter inserts (12 wells Transwell plates, Costar) at ca. 100,000 cells per filter (growth area $\pm 1 \text{ cm}^2$ containing 2.5 ml culture medium). The cells on the insert were cultured for 17 to 24 days at 37° C. in a humidified incubator containing 5% CO_2 in air. During this culture period the cells have been subjected to an enterocyte-like differentiation. Gentamycin was eliminated from the culture medium two days prior to the adhesion assays.

The adhesion assay was performed as follows. PBS was used as exposure medium. 25 μl of a bacterial suspension (2×10^9 cfu/ml) were added to 0.5 ml medium. The apical side of the Caco-2 monolayers was incubated with the bacterial suspensions for 1 hour at 37° C. After incubation, remaining fluid was removed and the cells were washed three times with 1 ml PBS. Subsequently, the Caco-2 monolayers were digested overnight with 1 ml 0.1M NaOH, 1% SDS. The lysate was mixed with 10 ml Hionic Fluor scintillation liquid and the radioactivity was measured by liquid scintillation counting using a LKB/Wallac scintillation counter. As a control, the radioactivity of the bacterial suspensions was measured. For each test group, the percentage of bacteria attached to the monolayers was calculated. All adhesion tests were performed in quadruple. In Table 4 the results of the bacterial adhesion test to Caco-2 cell lines are given. From the results can be concluded that the glucans and the fructans contribute to the adherence of *Lactobacillus reuteri* to Caco-2 cell lines. This could indicate that *Lactobacillus reuteri* strains producing EPS possess improved probiotic characteristics or that *Lactobacillus reuteri* and its polysaccharides could function as an excellent symbiotic.

TABLE 4

The results of the bacterial adhesion test to Caco-2 cell lines.		
Group (see Table 1)	0% of bacteria bound to the monolayer	
As	6.5	
Bs	5.7	
Cs	1.8	

TABLE 4-continued

The results of the bacterial adhesion test to Caco-2 cellines.	
Group (see Table 1)	0% of bacteria bound to the monolayer
D	2.3
E	0.9
F	1.3

DESCRIPTION OF THE FIGURES

FIG. 1: The nucleic acid (SEQ ID NO: 4) and deduced amino acid sequences (SEQ ID NOS 27 and 1) of the novel inulosucrase of *Lactobacillus reuteri*. Also encompassed within the figure is the comparison peptide (SEQ ID NO: 28). Furthermore, the designations and orientation (< for 3' to 5' and > for 5' to 3') of the primers and the restriction enzymes used for (inverse) PCR, are shown at the right hand side. Putative start codons (ATG, at positions 41 and 68) and stop codon (TAA, at position 2435) are shown in bold. The positions of the primers used for PCR are shown in bold/underlined. The *NheI* restriction sites (at positions 1154 and 2592) used for inverse PCR are underlined. The primers used and their exact positions in the inulosucrase sequence are shown in table 1. Starting at amino acid 690, the 20 PXX (residues 690-749 of SEQ ID NO: 1) repeats are underlined. At amino acid 755 the LPXTG (SEQ ID NO: 5) motif is underlined.

FIG. 2: Dendrogram of bacterial and plant fructosyltransferases. The horizontal distances are a measure for the

difference at the amino acid sequence level. 10% difference is indicated by the upper bar. Bootstrap values (in percentages) are given at the root of each tree. Fructosyltransferases of Gram positive bacteria are indicated in the lower half of the figure (*B. staerothermophilus* SurB; *B. amyloliquefaciens* SacB; *B. subtilis* SacB; *S. mutans* SacB; *L. reuteri* FtfA (inulosucrase); *S. salivarius* Ftf). Plant fructosyltransferases are indicated in the middle part of the figure (*Cynara scolymus* Ss-1 ft; *Allium cepa* F-6 gft; *Hordeum vulgare* Sf-6 ft). Fructosyltransferases of Gram negative bacteria are shown in the upper part of the figure (*Z. mobilis* LevU; *Z. mobilis* SucE2; *Z. mobilis* SacB; *E. amylovora* Lcs; *A. diazotrophicus* LsdA).

FIG. 3: The N-terminal (SEQ ID NO: 6) and three internal amino acid sequences (SEQ ID NOS 7-9) of the novel levansucrase of *Lactobacillus reuteri*.

FIG. 4: Parts of an alignment of the deduced amino acid sequences of some bacterial fructosyltransferase genes (SEQ ID NOS 29-40). Sequences in bold indicate the consensus sequences used to construct the degenerated primers 5 ftf, 6 ftfi and 12 ftfi. (*) indicates a position with a fully conserved amino acid residue. (:) indicates a position with a fully conserved 'strong' group: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW. (.) indicates a position with a fully conserved 'weaker' group: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, FVLIM, HFY. Groups are according to the Pam250 residue weight matrix described by Altschul et al. (1990) J. Mol. Biol. 215, 403-410.

FIG. 5: The strategy used for the isolation of the inulosucrase gene from *Lactobacillus reuteri* 121 chromosomal DNA.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 40

<210> SEQ ID NO 1

<211> LENGTH: 789

<212> TYPE: PRT

<213> ORGANISM: *Lactobacillus reuteri*

<400> SEQUENCE: 1

```

Met Tyr Lys Ser Gly Lys Asn Trp Ala Val Val Thr Leu Ser Thr Ala
 1           5           10           15

Ala Leu Val Phe Gly Ala Thr Thr Val Asn Ala Ser Ala Asp Thr Asn
 20           25           30

Ile Glu Asn Asn Asp Ser Ser Thr Val Gln Val Thr Thr Gly Asp Asn
 35           40           45

Asp Ile Ala Val Lys Ser Val Thr Leu Gly Ser Gly Gln Val Ser Ala
 50           55           60

Ala Ser Asp Thr Thr Ile Arg Thr Ser Ala Asn Ala Asn Ser Ala Ser
 65           70           75           80

Ser Ala Ala Asn Thr Gln Asn Ser Asn Ser Gln Val Ala Ser Ser Ala
 85           90           95

Ala Ile Thr Ser Ser Thr Ser Ser Ala Ala Ser Leu Asn Asn Thr Asp
100           105           110

Ser Lys Ala Ala Gln Glu Asn Thr Asn Thr Ala Lys Asn Asp Asp Thr
115           120           125

Gln Lys Ala Ala Pro Ala Asn Glu Ser Ser Glu Ala Lys Asn Glu Pro

```

-continued

130	135	140
Ala Val Asn Val Asn Asp Ser Ser Ala Ala Lys Asn Asp Asp Gln Gln		
145	150	155 160
Ser Ser Lys Lys Asn Thr Thr Ala Lys Leu Asn Lys Asp Ala Glu Asn		
	165	170 175
Val Val Lys Lys Ala Gly Ile Asp Pro Asn Ser Leu Thr Asp Asp Gln		
	180	185 190
Ile Lys Ala Leu Asn Lys Met Asn Phe Ser Lys Ala Ala Lys Ser Gly		
	195	200 205
Thr Gln Met Thr Tyr Asn Asp Phe Gln Lys Ile Ala Asp Thr Leu Ile		
	210	215 220
Lys Gln Asp Gly Arg Tyr Thr Val Pro Phe Phe Lys Ala Ser Glu Ile		
	225	230 235 240
Lys Asn Met Pro Ala Ala Thr Thr Lys Asp Ala Gln Thr Asn Thr Ile		
	245	250 255
Glu Pro Leu Asp Val Trp Asp Ser Trp Pro Val Gln Asp Val Arg Thr		
	260	265 270
Gly Gln Val Ala Asn Trp Asn Gly Tyr Gln Leu Val Ile Ala Met Met		
	275	280 285
Gly Ile Pro Asn Gln Asn Asp Asn His Ile Tyr Leu Leu Tyr Asn Lys		
	290	295 300
Tyr Gly Asp Asn Glu Leu Ser His Trp Lys Asn Val Gly Pro Ile Phe		
	305	310 315 320
Gly Tyr Asn Ser Thr Ala Val Ser Gln Glu Trp Ser Gly Ser Ala Val		
	325	330 335
Leu Asn Ser Asp Asn Ser Ile Gln Leu Phe Tyr Thr Arg Val Asp Thr		
	340	345 350
Ser Asp Asn Asn Thr Asn His Gln Lys Ile Ala Ser Ala Thr Leu Tyr		
	355	360 365
Leu Thr Asp Asn Asn Gly Asn Val Ser Leu Ala Gln Val Arg Asn Asp		
	370	375 380
Tyr Ile Val Phe Glu Gly Asp Gly Tyr Tyr Tyr Gln Thr Tyr Asp Gln		
	385	390 395 400
Trp Lys Ala Thr Asn Lys Gly Ala Asp Asn Ile Ala Met Arg Asp Ala		
	405	410 415
His Val Ile Glu Asp Gly Asn Gly Asp Arg Tyr Leu Val Phe Glu Ala		
	420	425 430
Ser Thr Gly Leu Glu Asn Tyr Gln Gly Glu Asp Gln Ile Tyr Asn Trp		
	435	440 445
Leu Asn Tyr Gly Gly Asp Asp Ala Phe Asn Ile Lys Ser Leu Phe Arg		
	450	455 460
Ile Leu Ser Asn Asp Asp Ile Lys Ser Arg Ala Thr Trp Ala Asn Ala		
	465	470 475 480
Ala Ile Gly Ile Leu Lys Leu Asn Lys Asp Glu Lys Asn Pro Lys Val		
	485	490 495
Ala Glu Leu Tyr Ser Pro Leu Ile Ser Ala Pro Met Val Ser Asp Glu		
	500	505 510
Ile Glu Arg Pro Asn Val Val Lys Leu Gly Asn Lys Tyr Tyr Leu Phe		
	515	520 525
Ala Ala Thr Arg Leu Asn Arg Gly Ser Asn Asp Asp Ala Trp Met Asn		
	530	535 540
Ala Asn Tyr Ala Val Gly Asp Asn Val Ala Met Val Gly Tyr Val Ala		
	545	550 555 560

-continued

Asp Ser Leu Thr Gly Ser Tyr Lys Pro Leu Asn Asp Ser Gly Val Val
 565 570 575
 Leu Thr Ala Ser Val Pro Ala Asn Trp Arg Thr Ala Thr Tyr Ser Tyr
 580 585 590
 Tyr Ala Val Pro Val Ala Gly Lys Asp Asp Gln Val Leu Val Thr Ser
 595 600 605
 Tyr Met Thr Asn Arg Asn Gly Val Ala Gly Lys Gly Met Asp Ser Thr
 610 615 620
 Trp Ala Pro Ser Phe Leu Leu Gln Ile Asn Pro Asp Asn Thr Thr Thr
 625 630 635 640
 Val Leu Ala Lys Met Thr Asn Gln Gly Asp Trp Ile Trp Asp Asp Ser
 645 650 655
 Ser Glu Asn Leu Asp Met Ile Gly Asp Leu Asp Ser Ala Ala Leu Pro
 660 665 670
 Gly Glu Arg Asp Lys Pro Val Asp Trp Asp Leu Ile Gly Tyr Gly Leu
 675 680 685
 Lys Pro His Asp Pro Ala Thr Pro Asn Asp Pro Glu Thr Pro Thr Thr
 690 695 700
 Pro Glu Thr Pro Glu Thr Pro Asn Thr Pro Lys Thr Pro Lys Thr Pro
 705 710 715 720
 Glu Asn Pro Gly Thr Pro Gln Thr Pro Asn Thr Pro Asn Thr Pro Glu
 725 730 735
 Ile Pro Leu Thr Pro Glu Thr Pro Lys Gln Pro Glu Thr Gln Thr Asn
 740 745 750
 Asn Arg Leu Pro Gln Thr Gly Asn Asn Ala Asn Lys Ala Met Ile Gly
 755 760 765
 Leu Gly Met Gly Thr Leu Leu Ser Met Phe Gly Leu Ala Glu Ile Asn
 770 775 780
 Lys Arg Arg Phe Asn
 785

<210> SEQ ID NO 2
 <211> LENGTH: 2367
 <212> TYPE: DNA
 <213> ORGANISM: Lactobacillus reuteri

<400> SEQUENCE: 2

atgtataaaa gcggtataaaa ttgggcagtc gttacactct cgactgctgc gctggatttt 60
 ggtgcaacaa ctgtaatatgc atccgcggac acaaatattg aaacaatga ttcttctact 120
 gtacaagtta caacaggatga taatgatatt gctgttaaaa gtgtgacact tggtagtggt 180
 caagttagtgc cagctagtga tacgactatt agaacttctg ctaatgcaaa tagtgcttct 240
 tctgccgcta atacacaaaa ttctaacagt caagtagcaa gttctgctgc aataacatca 300
 tctacaagtt ccgcagcttc attaaataac acagatagta aagcggctca agaaaatact 360
 aatacagcca aaatgatga cagcgaacaa gctgcaccag ctaacgaatc ttctgaagct 420
 aaaaatgaac cagctgtaaa cgtaaatgat tcttcagctg caaaaaatga tgatcaacaa 480
 tccagtataaa agaatactac cgctaagtta aacaaggatg ctgaaaaagt tgtaaaaaag 540
 gcgggaattg atcctaacag ttttaactgat gaccagatta aagcattaaa taagatgaac 600
 ttctcgaaag ctgcaaatgc tggtagacaa atgacttata atgatttcca aaagattgct 660
 gatacgtaa tcaacaaga tggtcggtag acagttccat tctttaaaag aagtgaatc 720
 aaaaatatgc ctgccgctac aactaaagat gcacaaacta atactattga acctttagat 780

-continued

gtatgggatt catggccagt tcaagatgtt cggacaggac aagttgctaa ttggaatggc	840
tatcaaacttg tcatcgcaat gatgggaatt ccaaaccaaa atgataatca tatctatctc	900
ttatataata agtatgggta taatgaatta agtcattgga agaattgtagg tccaattttt	960
ggctataatt ctaccgggtt ttcacaagaa tggtcaggat cagctgtttt gaacagtgat	1020
aaacttatcc aattatttta tacaagggtg gacacgtctg ataacaatac caatcatcaa	1080
aaaattgcta gcgctactct ttatttaact gataataatg gaaatgtatc actcgctcag	1140
gtacgaaatg actatattgt atttgaagggt gatggctatt actaccnaac ttatgatcaa	1200
tggaaagcta ctaacaaggg tgccgataat attgcaatgc gtgatgctca tgtaattgaa	1260
gatggtaatg gtgatcggtg ccttgttttt gaagcaagta ctggtttgga aaattatcaa	1320
ggcgaggacc aaatttataa ctgggttaaat tatggcggag atgacgcatt taatatcaag	1380
agcttattta gaattctttc caatgatgat attaagagtc gggcaacttg ggctaagtca	1440
gctatcggtg tcctcaaaat aaataaggac gaaaagaatc ctaaggtggc agagtatac	1500
tcaccattaa tttctgcacc aatggtaagc gatgaaattg agcgaccaa tgtagttaa	1560
ttaggtaata aatattactt atttgcgct acccgtttaa atcgaggag taatgatgat	1620
gcttgatga atgctaatta tgccgttggt gataatgttg caatggtcgg atatgttgct	1680
gatagtctaa ctggatctta taagccatta aatgattctg gagtagtctt gactgcttct	1740
gttctgcaa actggcggac agcaacttat tcatattatg ctgtcccggt tgccggaaaa	1800
gatgaccaag tattagttac ttcatatatg actaatagaa atggagtagc gggtaaagga	1860
atggattcaa ctgggcacc gagtttctta ctacaaatta acccgataa cacaactact	1920
gttttagcta aaatgactaa tcaaggggat tggatttggg atgattcaag cgaatctt	1980
gatatgattg gtgatttaga ctccgctgct ttacctggcg aacgtgataa acctgttgat	2040
tgggacttaa ttggttatgg attaaaaccg catgatcctg ctacaccaa tgatcctgaa	2100
acgccaacta caccagaac ccctgagaca cctaatactc ccaaaccacc aaagactcct	2160
gaaaatcctg ggacacctca aactcctaata acacctaata ctccggaat tcctttaact	2220
ccagaaacgc ctaagcaacc tgaaccccaa actaataatc gtttgccaca aactggaaat	2280
aatgccaata aagccatgat tggcctaggt atgggaacat tgcttagtat gtttggtctt	2340
gcagaaatta acaaacgtcg atttaac	2367

<210> SEQ ID NO 3

<211> LENGTH: 2394

<212> TYPE: DNA

<213> ORGANISM: *Lactobacillus reuteri*

<400> SEQUENCE: 3

atgctagaac gcaaggaa caaaaaatg tataaaagcg gtaaaaattg ggcagtcgtt	60
acactctcga ctgctgcgct ggtatttggg gcaacaactg taaatgcac cgcggacaca	120
aatattgaaa acaatgattc ttctactgta caagttacaa cagggtgata tgatattgct	180
gttaaaagtg tgacacttgg tagtggtcaa gttagtgcag ctagtgtac gactattaga	240
acttctgcta atgcaaatag tgcttcttct gccgctaata caaaaattc taacagtcaa	300
gtagcaagtt ctgctgcaat aacatcatct acaagttccg cagcttcatt aaataacaca	360
gtagtaaa cggtcaaga aaatactaata acagccaaaa atgatgacac gcaaaaagct	420
gcaccagcta acgaatcttc tgaagctaaa aatgaaccag ctgtaaacgt taatgattct	480

-continued

tcagctgcaa	aaaatgatga	tcaacaatcc	agtaaaaaga	atactaccgc	taagttaaac	540
aaggatgctg	aaaacgttgt	aaaaaaggcg	ggaattgata	ctaacagttt	aactgatgac	600
cagattaaag	cattaaataa	gatgaacttc	tcgaaagctg	caaagtctgg	tacacaaatg	660
acttataatg	atttccaaaa	gattgctgat	acgttaatca	aacaagatgg	tcggtacaca	720
gttcatttct	ttaaagcaag	tgaatcaaaa	aatatgcctg	ccgctacaac	taaagatgca	780
caactaata	ctattgaacc	tttagatgta	tgggattcat	ggccagtcca	agatgttcgg	840
acaggacaag	ttgctaattg	gaatggctat	caacttgtca	tcgcaatgat	gggaattcca	900
aacccaaatg	ataatcatat	ctatctctta	tataataagt	atggtgataa	tgaattaagt	960
cattggaaga	atgtaggccc	aatttttggc	tataattcta	ccgcggtttc	acaagaatgg	1020
tcaggatcag	ctgttttgaa	cagtgataac	tctatccaat	tattttatag	aagggtagac	1080
acgtctgata	acaataccaa	tcatacaaaa	attgctagcg	ctactcttta	tttaactgat	1140
aataatggaa	atgtatcact	cgctcaggta	cgaaatgact	atattgtatt	tgaaggatgat	1200
ggctattact	accaaactta	tgatcaatgg	aaagctacta	acaaaggatg	cataaatatt	1260
gcaatgcgtg	atgctcatgt	aattgaagat	ggtaaatggg	atcggtacct	tgtttttgaa	1320
gcaagtactg	gtttggaaaa	ttatcaagcg	gaggacccaa	tttataactg	gttaaatatt	1380
ggcggagatg	acgcatttaa	tatcaagagc	ttatttagaa	ttctttccaa	tgatgatatt	1440
aagatcggg	caacttgggc	taatgcagct	atcggtatcc	tcaaaactaa	taaggacgaa	1500
aagaatccta	aggatggcga	gttataacta	ccattaattt	ctgcaccaat	ggtaagcgat	1560
gaaattgagc	gaccaaattg	agttaaatta	ggtaataaat	attacttatt	tgccgctacc	1620
cgtttaaatc	gaggaaagta	tgatgatgct	tggatgaatg	ctaattatgc	cgttggatgat	1680
aatgttgcaa	tggtcggata	tggtgctgat	agtctaactg	gatcttataa	gccattaaat	1740
gattctggag	tagtcttgac	tgcttctgtt	cctgcaaact	ggcggacagc	aacttattca	1800
tattatgctg	tcoccggtgc	cggaagaagat	gaccaagtat	tagttacttc	atatatgact	1860
aatagaaatg	gagtagcggg	taaaggaaatg	gattcaactt	gggcaccgag	tttcttacta	1920
caaatatacc	cggataaacac	aactactgtt	ttagctaaaa	tgactaatca	aggggattgg	1980
atttgggatg	attcaagcga	aaatcttgat	atgattgggtg	atttagactc	cgctgcttta	2040
cctggcgaac	gtgataaac	tggtgattgg	gacttaattg	gttatggatt	aaaaccgcat	2100
gatcctgcta	caccaaatga	tcctgaaacg	ccaactacac	cagaaacccc	tgagacacct	2160
aatactccca	aaacacccaa	gactcctgaa	aatcctggga	cacctcaaac	tcctaataca	2220
cctaatactc	cggaaattcc	tttaactcca	gaaacgccta	agcaacctga	aacccaaact	2280
aataatcggt	tgccacaaac	tggaataaat	gccaataaag	ccatgattgg	cctaggtatg	2340
ggaacattgc	ttagtatggt	tggtcttgca	gaaattaaca	aacgtcgatt	taac	2394

<210> SEQ ID NO 4
 <211> LENGTH: 2592
 <212> TYPE: DNA
 <213> ORGANISM: Lactobacillus reuteri
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (1)..(51)
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (68)..(2434)

<400> SEQUENCE: 4

tac aat ggg gtg gcg gag gtg aag aaa cgg ggt tac ttc tat gct aga 48

-continued

Tyr	Asn	Gly	Val	Ala	Glu	Val	Lys	Lys	Arg	Gly	Tyr	Phe	Tyr	Ala	Arg
1				5					10					15	
acg	caaggaacat	aaaaaa	atg	tat	aaa	agc	ggt	aaa	aat	tggt	gca	gtc	ggt		100
Thr					Met	Tyr	Lys	Ser	Gly	Lys	Asn	Trp	Ala	Val	Val
							20					25			
aca	ctc	tcg	act	gct	gcg	ctg	gta	ttt	ggt	gca	aca	act	gta	aat	gca
Thr	Leu	Ser	Thr	Ala	Ala	Leu	Val	Phe	Gly	Ala	Thr	Thr	Val	Asn	Ala
	30					35					40				
tcc	gcg	gac	aca	aat	att	gaa	aac	aat	gat	tct	tct	act	gta	caa	gtt
Ser	Ala	Asp	Thr	Asn	Ile	Glu	Asn	Asn	Asp	Ser	Ser	Thr	Val	Gln	Val
	45				50				55					60	
aca	aca	ggt	gat	aat	gat	att	gct	gtt	aaa	agt	gtg	aca	ctt	ggt	agt
Thr	Thr	Gly	Asp	Asn	Asp	Ile	Ala	Val	Lys	Ser	Val	Thr	Leu	Gly	Ser
			65					70					75		
ggt	caa	gtt	agt	gca	gct	agt	gat	acg	act	att	aga	act	tct	gct	aat
Gly	Gln	Val	Ser	Ala	Ala	Ser	Asp	Thr	Thr	Ile	Arg	Thr	Ser	Ala	Asn
		80					85					90			
gca	aat	agt	gct	tct	tct	gcc	gct	aat	aca	caa	aat	tct	aac	agt	caa
Ala	Asn	Ser	Ala	Ser	Ser	Ala	Ala	Asn	Thr	Gln	Asn	Ser	Asn	Ser	Gln
		95				100					105				
gta	gca	agt	tct	gct	gca	ata	aca	tca	tct	aca	agt	tcc	gca	gct	tca
Val	Ala	Ser	Ser	Ala	Ala	Ile	Thr	Ser	Ser	Thr	Ser	Ser	Ala	Ala	Ser
	110					115					120				
tta	aat	aac	aca	gat	agt	aaa	gcg	gct	caa	gaa	aat	act	aat	aca	gcc
Leu	Asn	Asn	Thr	Asp	Ser	Lys	Ala	Ala	Gln	Glu	Asn	Thr	Asn	Thr	Ala
	125				130				135					140	
aaa	aat	gat	gac	acg	caa	aaa	gct	gca	cca	gct	aac	gaa	tct	tct	gaa
Lys	Asn	Asp	Asp	Thr	Gln	Lys	Ala	Ala	Pro	Ala	Asn	Glu	Ser	Ser	Glu
			145						150				155		
gct	aaa	aat	gaa	cca	gct	gta	aac	gtt	aat	gat	tct	tca	gct	gca	aaa
Ala	Lys	Asn	Glu	Pro	Ala	Val	Asn	Val	Asn	Asp	Ser	Ser	Ala	Ala	Lys
			160				165						170		
aat	gat	gat	caa	caa	tcc	agt	aaa	aag	aat	act	acc	gct	aag	tta	aac
Asn	Asp	Asp	Gln	Gln	Ser	Ser	Lys	Lys	Asn	Thr	Thr	Ala	Lys	Leu	Asn
		175				180						185			
aag	gat	gct	gaa	aac	gtt	gta	aaa	aag	gcg	gga	att	gat	cct	aac	agt
Lys	Asp	Ala	Glu	Asn	Val	Val	Lys	Lys	Ala	Gly	Ile	Asp	Pro	Asn	Ser
	190					195					200				
tta	act	gat	gac	cag	att	aaa	gca	tta	aat	aag	atg	aac	ttc	tcg	aaa
Leu	Thr	Asp	Asp	Gln	Ile	Lys	Ala	Leu	Asn	Lys	Met	Asn	Phe	Ser	Lys
	205				210					215				220	
gct	gca	aag	tct	ggt	aca	caa	atg	act	tat	aat	gat	ttc	caa	aag	att
Ala	Ala	Lys	Ser	Gly	Thr	Gln	Met	Thr	Tyr	Asn	Asp	Phe	Gln	Lys	Ile
			225						230				235		
gct	gat	acg	tta	atc	aaa	caa	gat	ggt	cggt	tac	aca	gtt	cca	ttc	ttt
Ala	Asp	Thr	Leu	Ile	Lys	Gln	Asp	Gly	Arg	Tyr	Thr	Val	Pro	Phe	Phe
		240					245					250			
aaa	gca	agt	gaa	atc	aaa	aat	atg	cct	gcc	gct	aca	act	aaa	gat	gca
Lys	Ala	Ser	Glu	Ile	Lys	Asn	Met	Pro	Ala	Ala	Thr	Thr	Lys	Asp	Ala
		255				260						265			
caa	act	aat	act	att	gaa	cct	tta	gat	gta	tggt	gat	tca	tggt	cca	gtt
Gln	Thr	Asn	Thr	Ile	Glu	Pro	Leu	Asp	Val	Trp	Asp	Ser	Trp	Pro	Val
		270				275					280				
caa	gat	gtt	cggt	aca	gga	caa	gtt	gct	aat	tggt	aat	ggc	tat	caa	ctt
Gln	Asp	Val	Arg	Thr	Gly	Gln	Val	Ala	Asn	Trp	Asn	Gly	Tyr	Gln	Leu
	285				290					295				300	
gtc	atc	gca	atg	atg	gga	att	cca	aac	caa	aat	gat	aat	cat	atc	tat
Val	Ile	Ala	Met	Met	Gly	Ile	Pro	Asn	Gln	Asn	Asp	Asn	His	Ile	Tyr
		305							310					315	

-continued

ctc tta tat aat aag tat ggt gat aat gaa tta agt cat tgg aag aat Leu Leu Tyr Asn Lys Tyr Gly Asp Asn Glu Leu Ser His Trp Lys Asn 320 325 330	1012
gta ggt cca att ttt ggc tat aat tct acc gcg gtt tca caa gaa tgg Val Gly Pro Ile Phe Gly Tyr Asn Ser Thr Ala Val Ser Gln Glu Trp 335 340 345	1060
tca gga tca gct gtt ttg aac agt gat aac tct atc caa tta ttt tat Ser Gly Ser Ala Val Leu Asn Ser Asp Asn Ser Ile Gln Leu Phe Tyr 350 355 360	1108
aca agg gta gac acg tct gat aac aat acc aat cat caa aaa att gct Thr Arg Val Asp Thr Ser Asp Asn Asn Thr Asn His Gln Lys Ile Ala 365 370 375 380	1156
agc gct act ctt tat tta act gat aat aat gga aat gta tca ctc gct Ser Ala Thr Leu Tyr Leu Thr Asp Asn Asn Gly Asn Val Ser Leu Ala 385 390 395	1204
cag gta cga aat gac tat att gta ttt gaa ggt gat ggc tat tac tac Gln Val Arg Asn Asp Tyr Ile Val Phe Glu Gly Asp Gly Tyr Tyr 400 405 410	1252
caa act tat gat caa tgg aaa gct act aac aaa ggt gcc gat aat att Gln Thr Tyr Asp Gln Trp Lys Ala Thr Asn Lys Gly Ala Asp Asn Ile 415 420 425	1300
gca atg cgt gat gct cat gta att gaa gat ggt aat ggt gat cgg tac Ala Met Arg Asp Ala His Val Ile Glu Asp Gly Asn Gly Asp Arg Tyr 430 435 440	1348
ctt gtt ttt gaa gca agt act ggt ttg gaa aat tat caa ggc gag gac Leu Val Phe Glu Ala Ser Thr Gly Leu Glu Asn Tyr Gln Gly Glu Asp 445 450 455 460	1396
caa att tat aac tgg tta aat tat ggc gga gat gac gca ttt aat atc Gln Ile Tyr Asn Trp Leu Asn Tyr Gly Gly Asp Asp Ala Phe Asn Ile 465 470 475	1444
aag agc tta ttt aga att ctt tcc aat gat gat att aag agt cgg gca Lys Ser Leu Phe Arg Ile Leu Ser Asn Asp Asp Ile Lys Ser Arg Ala 480 485 490	1492
act tgg gct aat gca gct atc ggt atc ctc aaa cta aat aag gac gaa Thr Trp Ala Asn Ala Ala Ile Gly Ile Leu Lys Leu Asn Lys Asp Glu 495 500 505	1540
aag aat cct aag gtg gca gag tta tac tca cca tta att tct gca cca Lys Asn Pro Lys Val Ala Glu Leu Tyr Ser Pro Leu Ile Ser Ala Pro 510 515 520	1588
atg gta agc gat gaa att gag cga cca aat gta gtt aaa tta ggt aat Met Val Ser Asp Glu Ile Glu Arg Pro Asn Val Val Lys Leu Gly Asn 525 530 535 540	1636
aaa tat tac tta ttt gcc gct acc cgt tta aat cga gga agt aat gat Lys Tyr Tyr Leu Phe Ala Ala Thr Arg Leu Asn Arg Gly Ser Asn Asp 545 550 555	1684
gat gct tgg atg aat gct aat tat gcc gtt ggt gat aat gtt gca atg Asp Ala Trp Met Asn Ala Asn Tyr Ala Val Gly Asp Asn Val Ala Met 560 565 570	1732
gtc gga tat gtt gct gat agt cta act gga tct tat aag cca tta aat Val Gly Tyr Val Ala Asp Ser Leu Thr Gly Ser Tyr Lys Pro Leu Asn 575 580 585	1780
gat tct gga gta gtc ttg act gct tct gtt cct gca aac tgg cgg aca Asp Ser Gly Val Val Leu Thr Ala Ser Val Pro Ala Asn Trp Arg Thr 590 595 600	1828
gca act tat tca tat tat gct gtc ccc gtt gcc gga aaa gat gac caa Ala Thr Tyr Ser Tyr Tyr Ala Val Pro Val Ala Gly Lys Asp Asp Gln 605 610 615 620	1876
gta tta gtt act tca tat atg act aat aga aat gga gta gcg ggt aaa Val Leu Val Thr Ser Tyr Met Thr Asn Arg Asn Gly Val Ala Gly Lys 625 630 635	1924

-continued

```

gga atg gat tca act tgg gca ccg agt ttc tta cta caa att aac ccg 1972
Gly Met Asp Ser Thr Trp Ala Pro Ser Phe Leu Leu Gln Ile Asn Pro
      640              645              650

gat aac aca act act gtt tta gct aaa atg act aat caa ggg gat tgg 2020
Asp Asn Thr Thr Val Leu Ala Lys Met Thr Asn Gln Gly Asp Trp
      655              660              665

att tgg gat gat tca agc gaa aat ctt gat atg att ggt gat tta gac 2068
Ile Trp Asp Asp Ser Ser Glu Asn Leu Asp Met Ile Gly Asp Leu Asp
      670              675              680

tcc gct gct tta cct ggc gaa cgt gat aaa cct gtt gat tgg gac tta 2116
Ser Ala Ala Leu Pro Gly Glu Arg Asp Lys Pro Val Asp Trp Asp Leu
      685              690              695              700

att ggt tat gga tta aaa ccg cat gat cct gct aca cca aat gat cct 2164
Ile Gly Tyr Gly Leu Lys Pro His Asp Pro Ala Thr Pro Asn Asp Pro
      705              710              715

gaa acg cca act aca cca gaa acc cct gag aca cct aat act ccc aaa 2212
Glu Thr Pro Thr Thr Pro Glu Thr Pro Glu Thr Pro Asn Thr Pro Lys
      720              725              730

aca cca aag act cct gaa aat cct ggg aca cct caa act cct aat aca 2260
Thr Pro Lys Thr Pro Glu Asn Pro Gly Thr Pro Gln Thr Pro Asn Thr
      735              740              745

cct aat act ccg gaa att cct tta act cca gaa acg cct aag caa cct 2308
Pro Asn Thr Pro Glu Ile Pro Leu Thr Pro Glu Thr Pro Lys Gln Pro
      750              755              760

gaa acc caa act aat aat cgt ttg cca caa act gga aat aat gcc aat 2356
Glu Thr Gln Thr Asn Asn Arg Leu Pro Gln Thr Gly Asn Asn Ala Asn
      765              770              775              780

aaa gcc atg att ggc cta ggt atg gga aca ttg ctt agt atg ttt ggt 2404
Lys Ala Met Ile Gly Leu Gly Met Gly Thr Leu Leu Ser Met Phe Gly
      785              790              795

ctt gca gaa att aac aaa cgt cga ttt aac taaatacttt aaaataaaac 2454
Leu Ala Glu Ile Asn Lys Arg Arg Phe Asn
      800              805

cgctaagcct taaattcagc ttaacgggtt tttattttta aagtttttat tgtaaaaaag 2514

cgaattatca ttaatactaa tgcaattggt gtaagacctt acgacagtag taacaatgaa 2574

tttgcccatc ttgtcgg 2592

```

```

<210> SEQ ID NO 5
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Lactobacillus reuteri
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (3)
<223> OTHER INFORMATION: Any amino acid

```

```

<400> SEQUENCE: 5

```

```

Leu Pro Xaa Thr Gly
 1              5

```

```

<210> SEQ ID NO 6
<211> LENGTH: 27
<212> TYPE: PRT
<213> ORGANISM: Lactobacillus reuteri

```

```

<400> SEQUENCE: 6

```

```

Ala Gln Val Glu Ser Asn Asn Tyr Asn Gly Val Ala Glu Val Asn Thr
 1              5              10              15

Glu Arg Gln Ala Asn Gly Gln Ile Gly Val Asp
      20              25

```

-continued

<210> SEQ ID NO 7
 <211> LENGTH: 16
 <212> TYPE: PRT
 <213> ORGANISM: Lactobacillus reuteri

<400> SEQUENCE: 7

Met Ala His Leu Asp Val Trp Asp Ser Trp Pro Val Gln Asp Pro Val
 1 5 10 15

<210> SEQ ID NO 8
 <211> LENGTH: 9
 <212> TYPE: PRT
 <213> ORGANISM: Lactobacillus reuteri

<400> SEQUENCE: 8

Asn Ala Gly Ser Ile Phe Gly Thr Lys
 1 5

<210> SEQ ID NO 9
 <211> LENGTH: 19
 <212> TYPE: PRT
 <213> ORGANISM: Lactobacillus reuteri

<400> SEQUENCE: 9

Val Glu Glu Val Tyr Ser Pro Lys Val Thr Leu Met Ala Ser Asp
 1 5 10 15

Glu Val Glu

<210> SEQ ID NO 10
 <211> LENGTH: 4634
 <212> TYPE: DNA
 <213> ORGANISM: Lactobacillus reuteri
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (1220)..(3598)
 <220> FEATURE:
 <221> NAME/KEY: RBS
 <222> LOCATION: (1205)..(1210)
 <220> FEATURE:
 <221> NAME/KEY: modified_base
 <222> LOCATION: (2702)..(2707)
 <223> OTHER INFORMATION: a, c, t, g, other or unknown
 <220> FEATURE:
 <221> NAME/KEY: modified_base
 <222> LOCATION: (3686)..(3698)
 <223> OTHER INFORMATION: a, c, t, g, other or unknown

<400> SEQUENCE: 10

gttaacaaag acaaaatttt atataattct tcaaatataa tttccactg taagaacata 60
 aatgggtacc tgtttgatgg gaataatata tttgtaacta accggccggc acctctttct 120
 aatgtgccta ggatgcataa tggatgtaaa ttactagatg gcggttttta tacattaacc 180
 tcgcaggaga gaaaagaagc aattagtaag gatccatag cagataaatt tattaggcct 240
 tatttaggtg ctaaaaattt cattcatgga actgctaggt actgtatttg gttaaaggac 300
 gcaaacccga aagatatcca tcaatcgcca tttatactgg atagaatcaa taaagtagcg 360
 gaattcagat cgcagcaaaa aagtaagat acacaaaaat atgcaaaacg gcccatgcta 420
 acaacacgac ttgcctatta tagccacgat gtacatacgg atatgctgat agtacctgca 480
 acatcatcgc aacgtagaga atatcttcca attggatag tttcagaaaa gaattatttg 540
 tcttattcac taatgctaata ccccaatgct agtaatttta atttcggtat tctagaatct 600
 aaagttcact atatttggtt aaaaaacttt tgcggtcggt tgaagtcoga ttatcggtat 660

-continued

tcaaacacta ttatttataa taatttcctt tggccgactg ttggtgacaa gccaggamca	720
acaccatctc tgacactcgc tcaaggtata ttaataactc gcaagctcta tccagacagc	780
tcactggctg atctttatga tccactaaca atgccragtt gaactcgtaa agctcatgaa	840
gccaatgata aagctgttct taaagcatat ggattgagcc ctaaagctac tgagcaagaa	900
atcgtagaac atctatttaa gatgtatgaa aaactgacta aaggtgaaag ataactttgt	960
aaaaccaata ttttataaag acagtaaatg ttaatttgat aaaaacatat atttaataaa	1020
caaaagtgat ataataaagt agttctttgt attacaaaat acatttaata tctctcagca	1080
ttttgcatac tgggagattt tttattgaca aattgtttga aagtgttat gatgaaaccg	1140
tgtagaaact aattcaattt gataaacgtt agacatttct gaggaggaaag tcattttgga	1200
gtacaaagaa cataagaaa atg tat aaa gtc ggc aag aat tgg gcc gtt gct	1252
Met Tyr Lys Val Gly Lys Asn Trp Ala Val Ala	
1 5 10	
aca ttg gta tca gct tca att tta atg gga ggg gtt gta acc gct cat	1300
Thr Leu Val Ser Ala Ser Ile Leu Met Gly Gly Val Val Thr Ala His	
15 20 25	
gct gat caa gta gaa agt aac aat tac aac ggt gtt gct gaa gtt aat	1348
Ala Asp Gln Val Glu Ser Asn Asn Tyr Asn Gly Val Ala Glu Val Asn	
30 35 40	
act gaa cgt caa gct aat ggt caa att ggc gta gat gga aaa att att	1396
Thr Glu Arg Gln Ala Asn Gly Gln Ile Gly Val Asp Gly Lys Ile Ile	
45 50 55	
agt gct aac agt aat aca acc agt ggc tcg aca aat caa gaa tca tct	1444
Ser Ala Asn Ser Asn Thr Thr Ser Gly Ser Thr Asn Gln Glu Ser Ser	
60 65 70 75	
gct act aac aat act gaa aat gct gtt gtt aat gaa agc aaa aat act	1492
Ala Thr Asn Asn Thr Glu Asn Ala Val Val Asn Glu Ser Lys Asn Thr	
80 85 90	
aac aat act gaa aat gct gtt gtt aat gaa aac aaa aat act aac aat	1540
Asn Asn Thr Glu Asn Ala Val Val Asn Glu Asn Lys Asn Thr Asn Asn	
95 100 105	
act gaa aat gct gtt gtt aat gaa aac aaa aat act aac aac aca gaa	1588
Thr Glu Asn Ala Val Val Asn Glu Asn Lys Asn Thr Asn Asn Thr Glu	
110 115 120	
aac gat aat agt caa tta aag tta act aat aat gaa caa cca tca gcc	1636
Asn Asp Asn Ser Gln Leu Lys Leu Thr Asn Asn Glu Gln Pro Ser Ala	
125 130 135	
gct act caa gca aac ttg aag aag cta aat cct caa gct gct aag gct	1684
Ala Thr Gln Ala Asn Leu Lys Lys Leu Asn Pro Gln Ala Ala Lys Ala	
140 145 150 155	
ggt caa aat gcc aag att gat gcc ggt agt tta aca gat gat caa att	1732
Val Gln Asn Ala Lys Ile Asp Ala Gly Ser Leu Thr Asp Asp Gln Ile	
160 165 170	
aat gaa tta aat aag att aac ttc tct aag tct gct gaa aag ggt gca	1780
Asn Glu Leu Asn Lys Ile Asn Phe Ser Lys Ser Ala Glu Lys Gly Ala	
175 180 185	
aaa ttg acc ttt aag gac tta gag ggg att ggt aat gct att gtt aag	1828
Lys Leu Thr Phe Lys Asp Leu Glu Gly Ile Gly Asn Ala Ile Val Lys	
190 195 200	
caa gat cca caa tat gct att cct tat tct aat gct aag gaa atc aag	1876
Gln Asp Pro Gln Tyr Ala Ile Pro Tyr Ser Asn Ala Lys Glu Ile Lys	
205 210 215	
aat atg cct gca aca tac act gta gat gcc caa aca ggt aag atg gct	1924
Asn Met Pro Ala Thr Tyr Thr Val Asp Ala Gln Thr Gly Lys Met Ala	
220 225 230 235	

-continued

cat ctt gat gtc tgg gac tct tgg cca gta caa gat cct gtc aca ggt	1972
His Leu Asp Val Trp Asp Ser Trp Pro Val Gln Asp Pro Val Thr Gly	
240 245 250	
tat gta tct aat tac atg ggt tat caa cta gtt att gct atg atg ggt	2020
Tyr Val Ser Asn Tyr Met Gly Tyr Gln Leu Val Ile Ala Met Met Gly	
255 260 265	
att cca aat tgc cca act gga gat aat cat atc tat ctt ctt tac aac	2068
Ile Pro Asn Ser Pro Thr Gly Asp Asn His Ile Tyr Leu Leu Tyr Asn	
270 275 280	
aag tat ggt gat aat gac ttt tct cat tgg cgc aat gca ggt tca atc	2116
Lys Tyr Gly Asp Asn Asp Phe Ser His Trp Arg Asn Ala Gly Ser Ile	
285 290 295	
ttt gga act aaa gaa aca aat gtg ttc caa gaa tgg tca ggt tca gct	2164
Phe Gly Thr Lys Glu Thr Asn Val Phe Gln Glu Trp Ser Gly Ser Ala	
300 305 310 315	
att gta aat gat gat ggt aca att caa cta ttt ttc acc tca aat gat	2212
Ile Val Asn Asp Asp Gly Thr Ile Gln Leu Phe Phe Thr Ser Asn Asp	
320 325 330	
acg tct gat tac aag ttg aat gat caa cgc ctt gct acc gca aca tta	2260
Thr Ser Asp Tyr Lys Leu Asn Asp Gln Arg Leu Ala Thr Ala Thr Leu	
335 340 345	
aac ctt aat gtt gat gat aac ggt gtt tca atc aag agt gtt gat aat	2308
Asn Leu Asn Val Asp Asp Asn Gly Val Ser Ile Lys Ser Val Asp Asn	
350 355 360	
tat caa gtt ttg ttt gaa ggt gat gga ttt cac tac caa act tat gaa	2356
Tyr Gln Val Leu Phe Glu Gly Asp Gly Phe His Tyr Gln Thr Tyr Glu	
365 370 375	
caa ttc gca aac ggc aaa gat cgt gaa aat gat gat tac tgc tta cgt	2404
Gln Phe Ala Asn Gly Lys Asp Arg Glu Asn Asp Asp Tyr Cys Leu Arg	
380 385 390 395	
gac cca cac gtt gtt caa tta gaa aat ggt gat cgt tat ctt gta ttc	2452
Asp Pro His Val Val Gln Leu Glu Asn Gly Asp Arg Tyr Leu Val Phe	
400 405 410	
gaa gct aat act ggg aca gaa gat tac caa agt gac gac caa att tat	2500
Glu Ala Asn Thr Gly Thr Glu Asp Tyr Gln Ser Asp Asp Gln Ile Tyr	
415 420 425	
aat tgg gct aac tat ggt ggc gat gat gcc ttc aat att aag agt tcc	2548
Asn Trp Ala Asn Tyr Gly Gly Asp Asp Ala Phe Asn Ile Lys Ser Ser	
430 435 440	
ttc aag ctt ttg aat aat aag aag gat cgt gaa ttg gct ggt tta gct	2596
Phe Lys Leu Leu Asn Asn Lys Lys Asp Arg Glu Leu Ala Gly Leu Ala	
445 450 455	
aat ggt gca ctt ggt atc tta aag ctc act aac aat caa agt aag cca	2644
Asn Gly Ala Leu Gly Ile Leu Lys Leu Thr Asn Asn Gln Ser Lys Pro	
460 465 470 475	
aag gtt gaa gaa gta tac tca cca ttg gta tct act ttg atg gct tgc	2692
Lys Val Glu Glu Val Tyr Ser Pro Leu Val Ser Thr Leu Met Ala Cys	
480 485 490	
gat gag gta nnn nnn aag ctt ggt gat aag tat tat ctc ttc tcc gta	2740
Asp Glu Val Xaa Xaa Lys Leu Gly Asp Lys Tyr Tyr Leu Phe Ser Val	
495 500 505	
act cgt gta agt cgt ggt tcc gat cgt gaa tta acc gct aag gat aac	2788
Thr Arg Val Ser Arg Gly Ser Asp Arg Glu Leu Thr Ala Lys Asp Asn	
510 515 520	
aca atc gtt ggt gat aac gtt gct atg att ggt tac gtt tcc gat agc	2836
Thr Ile Val Gly Asp Asn Val Ala Met Ile Gly Tyr Val Ser Asp Ser	
525 530 535	
tta atg ggt aag tac aag cca tta aat aac tca ggt gtc gta tta act	2884
Leu Met Gly Lys Tyr Lys Pro Leu Asn Asn Ser Gly Val Val Leu Thr	
540 545 550 555	

-continued

gca tca gta cct gca aac tgg cgt act gct act tat tcc tac tat gca Ala Ser Val Pro Ala Asn Trp Arg Thr Ala Thr Tyr Ser Tyr Tyr Ala 560 565 570	2932
gta cct gta gct ggt cat cct gat caa gta tta att act tct tac atg Val Pro Val Ala Gly His Pro Asp Gln Val Leu Ile Thr Ser Tyr Met 575 580 585	2980
agt aac aag gac ttt gct tca ggt gaa gga aac tat gca act tgg gca Ser Asn Lys Asp Phe Ala Ser Gly Glu Gly Asn Tyr Ala Thr Trp Ala 590 595 600	3028
cca agt ttc tta gta caa atc aat cca gat gac acg aca act gta tta Pro Ser Phe Leu Val Gln Ile Asn Pro Asp Asp Thr Thr Thr Val Leu 605 610 615	3076
gca cgt gca act aac caa ggt gac tgg gtg tgg gac gac tct agt cgg Ala Arg Ala Thr Asn Gln Gly Asp Trp Val Trp Asp Asp Ser Ser Arg 620 625 630 635	3124
aac gat aat atg ctc ggt gtt ctt aaa gaa ggt gca gct aac agt gcc Asn Asp Asn Met Leu Gly Val Leu Lys Glu Gly Ala Ala Asn Ser Ala 640 645 650	3172
gcc tta cca ggt gaa tgg ggt aag cca gtt gac tgg agt ttg att aac Ala Leu Pro Gly Glu Trp Gly Lys Pro Val Asp Trp Ser Leu Ile Asn 655 660 665	3220
aga agt cct ggc tta ggc tta aag cct cat caa cca gtt caa cca aag Arg Ser Pro Gly Leu Gly Leu Lys Pro His Gln Pro Val Gln Pro Lys 670 675 680	3268
att gat caa cct gat caa caa cct tct ggt caa aac act aag aat gtc Ile Asp Gln Pro Asp Gln Gln Pro Ser Gly Gln Asn Thr Lys Asn Val 685 690 695	3316
aca cca ggt aat ggt gat aag cct gct ggt aag gca act cct gat aac Thr Pro Gly Asn Gly Asp Lys Pro Ala Gly Lys Ala Thr Pro Asp Asn 700 705 710 715	3364
act aat att gat cca agt gca caa cct tct ggt caa aac act aat att Thr Asn Ile Asp Pro Ser Ala Gln Pro Ser Gly Gln Asn Thr Asn Ile 720 725 730	3412
gat cca agt gca caa met tct ggt caa aac act aag aat gtc aca cca Asp Pro Ser Ala Gln Xaa Ser Gly Gln Asn Thr Lys Asn Val Thr Pro 735 740 745	3460
ggt aat gag aaa caa ggt aag aat acc gat gca aaa caa tta cca caa Gly Asn Glu Lys Gln Gly Lys Asn Thr Asp Ala Lys Gln Leu Pro Gln 750 755 760	3508
aca ggt aat aag tct ggt tta gca gga ctt tac gct ggt tca tta ctt Thr Gly Asn Lys Ser Gly Leu Ala Gly Leu Tyr Ala Gly Ser Leu Leu 765 770 775	3556
gcc ttg ttt gga ttg gca gca att gaa aag cgt cac gct taa Ala Leu Phe Gly Leu Ala Ala Ile Glu Lys Arg His Ala 780 785 790	3598
tagagtaaaa aaacatcctc cactcaagtt acaagtagga taatatgtat tatttctacg	3658
cytagtcaag aggrattact ggacatannn nnnnnnnnnn tccagttacc aagtggaata	3718
tagtattatt ccacgctagt caggaggatt actgacatta ttggctacat ggccggtagt	3778
cctcttttct tttgtgacga attgtcaaac caagtgcac ggtttctcaa aaaacacctc	3838
atatgggggtt tcataattta acacttttcg aggacggcgg ttcagctgat gttggcagaa	3898
actgacgtcc ttatctgtat aatcatcaat attagccctt ttaggaaagt attccctaatt	3958
tagaccattg gtattttcat tgggtcctct ttcctctggt gaatagggat ctggccaata	4018
gatagctact cctaaacgtc ctgcaatc attcaagcca agaaattcac gcccatgac	4078
tggagtcaat gaatggacaa attctttagg aatagaccct aagagatcaa ttaagccctg	4138

-continued

```

atatttgaat tcggagaagg ggagttgtcc aacaattgcc gttataatac cagggttaat 4198
acggccctgg gcctctacgg taatattgta tttttggctc agatcagtga tagaaaccca 4258
cagatttagc ttgccggtgg agtgctgctt gaagtcttca attacttcgt taccatgttt 4318
gattgctaata ctgatgtgtc gttgttggg ttagtaggac atcataccac ctcctcataa 4378
aataaggtat aacaggaatt tctgtacta tatgatcctt ccaatataat aatattaggc 4438
cgataagaaa tgaccagcta ccatttcttg atgcttagtg aatataatcg gatgatacgt 4498
caccctcaa caatccaatt tcacggaggt gagtaatcat gccgagagct aggaatgatt 4558
ggaggaacga acacgggtcca tgcggcagtg gctatttga ttttagccaa agcagcgtaa 4618
ctgcttgcaa aagctt 4634

```

```

<210> SEQ ID NO 11
<211> LENGTH: 792
<212> TYPE: PRT
<213> ORGANISM: Lactobacillus reuteri
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (495)..(496)
<223> OTHER INFORMATION: Any amino acid
<220> FEATURE:
<221> NAME/KEY: MOD_RES
<222> LOCATION: (737)
<223> OTHER INFORMATION: Thr or Pro

```

```

<400> SEQUENCE: 11

```

```

Met Tyr Lys Val Gly Lys Asn Trp Ala Val Ala Thr Leu Val Ser Ala
 1          5          10          15
Ser Ile Leu Met Gly Gly Val Val Thr Ala His Ala Asp Gln Val Glu
 20          25          30
Ser Asn Asn Tyr Asn Gly Val Ala Glu Val Asn Thr Glu Arg Gln Ala
 35          40          45
Asn Gly Gln Ile Gly Val Asp Gly Lys Ile Ile Ser Ala Asn Ser Asn
 50          55          60
Thr Thr Ser Gly Ser Thr Asn Gln Glu Ser Ser Ala Thr Asn Asn Thr
 65          70          75          80
Glu Asn Ala Val Val Asn Glu Ser Lys Asn Thr Asn Asn Thr Glu Asn
 85          90          95
Ala Val Val Asn Glu Asn Lys Asn Thr Asn Asn Thr Glu Asn Ala Val
100          105          110
Val Asn Glu Asn Lys Asn Thr Asn Asn Thr Glu Asn Asp Asn Ser Gln
115          120          125
Leu Lys Leu Thr Asn Asn Glu Gln Pro Ser Ala Ala Thr Gln Ala Asn
130          135          140
Leu Lys Lys Leu Asn Pro Gln Ala Ala Lys Ala Val Gln Asn Ala Lys
145          150          155          160
Ile Asp Ala Gly Ser Leu Thr Asp Asp Gln Ile Asn Glu Leu Asn Lys
165          170          175
Ile Asn Phe Ser Lys Ser Ala Glu Lys Gly Ala Lys Leu Thr Phe Lys
180          185          190
Asp Leu Glu Gly Ile Gly Asn Ala Ile Val Lys Gln Asp Pro Gln Tyr
195          200          205
Ala Ile Pro Tyr Ser Asn Ala Lys Glu Ile Lys Asn Met Pro Ala Thr
210          215          220
Tyr Thr Val Asp Ala Gln Thr Gly Lys Met Ala His Leu Asp Val Trp
225          230          235          240

```

-continued

Asp	Ser	Trp	Pro	Val	Gln	Asp	Pro	Val	Thr	Gly	Tyr	Val	Ser	Asn	Tyr
				245					250					255	
Met	Gly	Tyr	Gln	Leu	Val	Ile	Ala	Met	Met	Gly	Ile	Pro	Asn	Ser	Pro
			260					265					270		
Thr	Gly	Asp	Asn	His	Ile	Tyr	Leu	Leu	Tyr	Asn	Lys	Tyr	Gly	Asp	Asn
		275					280					285			
Asp	Phe	Ser	His	Trp	Arg	Asn	Ala	Gly	Ser	Ile	Phe	Gly	Thr	Lys	Glu
	290					295					300				
Thr	Asn	Val	Phe	Gln	Glu	Trp	Ser	Gly	Ser	Ala	Ile	Val	Asn	Asp	Asp
	305				310					315				320	
Gly	Thr	Ile	Gln	Leu	Phe	Phe	Thr	Ser	Asn	Asp	Thr	Ser	Asp	Tyr	Lys
			325					330						335	
Leu	Asn	Asp	Gln	Arg	Leu	Ala	Thr	Ala	Thr	Leu	Asn	Leu	Asn	Val	Asp
		340					345						350		
Asp	Asn	Gly	Val	Ser	Ile	Lys	Ser	Val	Asp	Asn	Tyr	Gln	Val	Leu	Phe
		355					360					365			
Glu	Gly	Asp	Gly	Phe	His	Tyr	Gln	Thr	Tyr	Glu	Gln	Phe	Ala	Asn	Gly
	370					375					380				
Lys	Asp	Arg	Glu	Asn	Asp	Asp	Tyr	Cys	Leu	Arg	Asp	Pro	His	Val	Val
	385				390				395					400	
Gln	Leu	Glu	Asn	Gly	Asp	Arg	Tyr	Leu	Val	Phe	Glu	Ala	Asn	Thr	Gly
			405					410						415	
Thr	Glu	Asp	Tyr	Gln	Ser	Asp	Asp	Gln	Ile	Tyr	Asn	Trp	Ala	Asn	Tyr
	420						425						430		
Gly	Gly	Asp	Asp	Ala	Phe	Asn	Ile	Lys	Ser	Ser	Phe	Lys	Leu	Leu	Asn
	435					440					445				
Asn	Lys	Lys	Asp	Arg	Glu	Leu	Ala	Gly	Leu	Ala	Asn	Gly	Ala	Leu	Gly
	450				455				460						
Ile	Leu	Lys	Leu	Thr	Asn	Asn	Gln	Ser	Lys	Pro	Lys	Val	Glu	Glu	Val
	465			470					475					480	
Tyr	Ser	Pro	Leu	Val	Ser	Thr	Leu	Met	Ala	Cys	Asp	Glu	Val	Xaa	Xaa
			485					490						495	
Lys	Leu	Gly	Asp	Lys	Tyr	Tyr	Leu	Phe	Ser	Val	Thr	Arg	Val	Ser	Arg
		500					505						510		
Gly	Ser	Asp	Arg	Glu	Leu	Thr	Ala	Lys	Asp	Asn	Thr	Ile	Val	Gly	Asp
		515					520					525			
Asn	Val	Ala	Met	Ile	Gly	Tyr	Val	Ser	Asp	Ser	Leu	Met	Gly	Lys	Tyr
	530				535						540				
Lys	Pro	Leu	Asn	Asn	Ser	Gly	Val	Val	Leu	Thr	Ala	Ser	Val	Pro	Ala
	545			550					555					560	
Asn	Trp	Arg	Thr	Ala	Thr	Tyr	Ser	Tyr	Tyr	Ala	Val	Pro	Val	Ala	Gly
		565					570							575	
His	Pro	Asp	Gln	Val	Leu	Ile	Thr	Ser	Tyr	Met	Ser	Asn	Lys	Asp	Phe
		580				585						590			
Ala	Ser	Gly	Glu	Gly	Asn	Tyr	Ala	Thr	Trp	Ala	Pro	Ser	Phe	Leu	Val
		595				600					605				
Gln	Ile	Asn	Pro	Asp	Asp	Thr	Thr	Thr	Val	Leu	Ala	Arg	Ala	Thr	Asn
	610				615						620				
Gln	Gly	Asp	Trp	Val	Trp	Asp	Asp	Ser	Ser	Arg	Asn	Asp	Asn	Met	Leu
	625				630					635				640	
Gly	Val	Leu	Lys	Glu	Gly	Ala	Ala	Asn	Ser	Ala	Ala	Leu	Pro	Gly	Glu
		645						650					655		
Trp	Gly	Lys	Pro	Val	Asp	Trp	Ser	Leu	Ile	Asn	Arg	Ser	Pro	Gly	Leu

-continued

660	665	670
Gly Leu Lys Pro His Gln Pro Val Gln Pro Lys Ile Asp Gln Pro Asp		
675	680	685
Gln Gln Pro Ser Gly Gln Asn Thr Lys Asn Val Thr Pro Gly Asn Gly		
690	695	700
Asp Lys Pro Ala Gly Lys Ala Thr Pro Asp Asn Thr Asn Ile Asp Pro		
705	710	715
Ser Ala Gln Pro Ser Gly Gln Asn Thr Asn Ile Asp Pro Ser Ala Gln		
725	730	735
Xaa Ser Gly Gln Asn Thr Lys Asn Val Thr Pro Gly Asn Glu Lys Gln		
740	745	750
Gly Lys Asn Thr Asp Ala Lys Gln Leu Pro Gln Thr Gly Asn Lys Ser		
755	760	765
Gly Leu Ala Gly Leu Tyr Ala Gly Ser Leu Leu Ala Leu Phe Gly Leu		
770	775	780
Ala Ala Ile Glu Lys Arg His Ala		
785	790	
<210> SEQ ID NO 12		
<211> LENGTH: 24		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: Description of Artificial Sequence: Primer		
<400> SEQUENCE: 12		
ctgataataa tggaaatgta tcac		
		24
<210> SEQ ID NO 13		
<211> LENGTH: 26		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: Description of Artificial Sequence: Primer		
<400> SEQUENCE: 13		
catgatcata agtttggtag taatag		
		26
<210> SEQ ID NO 14		
<211> LENGTH: 24		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: Description of Artificial Sequence: Primer		
<400> SEQUENCE: 14		
gtgatacatt tccattatta tcag		
		24
<210> SEQ ID NO 15		
<211> LENGTH: 26		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: Description of Artificial Sequence: Primer		
<400> SEQUENCE: 15		
ctattactac caaacttatg atcatg		
		26
<210> SEQ ID NO 16		
<211> LENGTH: 38		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		

-continued

<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Primer

<400> SEQUENCE: 16
ccatggccat ggtagaacgc aaggaacata aaaaaatg 38

<210> SEQ ID NO 17
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Primer

<400> SEQUENCE: 17
agatctagat ctgttaaatc gacgtttgtt aatttctg 38

<210> SEQ ID NO 18
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Primer
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (6)
<223> OTHER INFORMATION: a, c, t, g, other or unknown
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (15)
<223> OTHER INFORMATION: a, c, t, g, other or unknown

<400> SEQUENCE: 18
gaygtntggg aywsntgggc c 21

<210> SEQ ID NO 19
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (3)
<223> OTHER INFORMATION: a, c, t, g, other or unknown
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (6)
<223> OTHER INFORMATION: a, c, t, g, other or unknown
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)
<223> OTHER INFORMATION: a, c, t, g, other or unknown
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (12)
<223> OTHER INFORMATION: a, c, t, g, other or unknown
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Primer

<400> SEQUENCE: 19
gtngcswnc cswccayts ytg 23

<210> SEQ ID NO 20
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Primer

<400> SEQUENCE: 20
gaatgtaggt ccaatttttg gc 22

-continued

<210> SEQ ID NO 21
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Primer

<400> SEQUENCE: 21

cctgtccgaa catcttgaac tg

22

<210> SEQ ID NO 22
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Primer
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (6)
<223> OTHER INFORMATION: a, c, t, g, other or unknown
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)
<223> OTHER INFORMATION: a, c, t, g, other or unknown
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (12)
<223> OTHER INFORMATION: a, c, t, g, other or unknown
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (18)
<223> OTHER INFORMATION: a, c, t, g, other or unknown
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (21)
<223> OTHER INFORMATION: a, c, t, g, other or unknown

<400> SEQUENCE: 22

arraanswng gngcvmangt nsw

23

<210> SEQ ID NO 23
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Primer
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (9)
<223> OTHER INFORMATION: a, c, t, g, other or unknown
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (11)
<223> OTHER INFORMATION: a, c, t, g, other or unknown
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (15)
<223> OTHER INFORMATION: a, c, t, g, other or unknown
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (21)
<223> OTHER INFORMATION: a, c, t, g, other or unknown

<400> SEQUENCE: 23

tayaayggng tngcngargt naa

23

<210> SEQ ID NO 24
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Primer

-continued

<400> SEQUENCE: 24

ccgaccatct tgtttgatta ac

22

<210> SEQ ID NO 25

<211> LENGTH: 24

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Primer

<400> SEQUENCE: 25

aaytataayg gygttgcryg aagt

24

<210> SEQ ID NO 26

<211> LENGTH: 21

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Primer

<220> FEATURE:

<221> NAME/KEY: modified_base

<222> LOCATION: (9)

<223> OTHER INFORMATION: a, c, t, g, other or unknown

<400> SEQUENCE: 26

taccgnwsnc tacttcaact t

21

<210> SEQ ID NO 27

<211> LENGTH: 17

<212> TYPE: PRT

<213> ORGANISM: Lactobacillus reuteri

<400> SEQUENCE: 27

Tyr Asn Gly Val Ala Glu Val Lys Lys Arg Gly Tyr Phe Tyr Ala Arg
1 5 10 15

Thr

<210> SEQ ID NO 28

<211> LENGTH: 17

<212> TYPE: PRT

<213> ORGANISM: Lactobacillus reuteri

<400> SEQUENCE: 28

Tyr Asn Gly Val Ala Glu Val Asn Thr Glu Arg Gln Ala Asn Gly Gly
1 5 10 15

Ile

<210> SEQ ID NO 29

<211> LENGTH: 14

<212> TYPE: PRT

<213> ORGANISM: Bacillus amyloliquefaciens

<400> SEQUENCE: 29

Gly Leu Asp Val Trp Asp Ser Trp Pro Leu Gln Asn Ala Asp
1 5 10

<210> SEQ ID NO 30

<211> LENGTH: 14

<212> TYPE: PRT

<213> ORGANISM: Bacillus subtilis

<400> SEQUENCE: 30

Gly Leu Asp Val Trp Asp Ser Trp Pro Leu Gln Asn Ala Asp

-continued

1	5	10
---	---	----

<210> SEQ ID NO 31
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Streptococcus mutans
<400> SEQUENCE: 31
Asp Leu Asp Val Trp Asp Ser Trp Pro Val Gln Asp Ala Lys
1 5 10

<210> SEQ ID NO 32
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Streptococcus salivarius
<400> SEQUENCE: 32
Glu Ile Asp Val Trp Asp Ser Trp Pro Val Gln Asp Ala Lys
1 5 10

<210> SEQ ID NO 33
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Bacillus amyloliquefaciens
<400> SEQUENCE: 33
Gln Thr Gln Glu Trp Ser Gly Ser Ala Thr Phe Thr Ser Asp Gly Lys
1 5 10 15

<210> SEQ ID NO 34
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Bacillus subtilis
<400> SEQUENCE: 34
Gln Thr Gln Glu Trp Ser Gly Ser Ala Thr Phe Thr Ser Asp Gly Lys
1 5 10 15

<210> SEQ ID NO 35
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Streptococcus mutans
<400> SEQUENCE: 35
Leu Thr Gln Glu Trp Ser Gly Ser Ala Thr Val Asn Glu Asp Gly Ser
1 5 10 15

<210> SEQ ID NO 36
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Streptococcus salivarius
<400> SEQUENCE: 36
Asp Asp Gln Gln Trp Ser Gly Ser Ala Thr Val Asn Ser Asp Gly Ser
1 5 10 15

<210> SEQ ID NO 37
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Bacillus amyloliquefaciens
<400> SEQUENCE: 37
Lys Ala Thr Phe Gly Pro Ser Phe Leu Met Asn
1 5 10

-continued

<210> SEQ ID NO 38
 <211> LENGTH: 11
 <212> TYPE: PRT
 <213> ORGANISM: *Bacillus subtilis*

<400> SEQUENCE: 38

Gln Ser Thr Phe Ala Pro Ser Phe Leu Leu Asn
 1 5 10

<210> SEQ ID NO 39
 <211> LENGTH: 11
 <212> TYPE: PRT
 <213> ORGANISM: *Streptococcus mutans*

<400> SEQUENCE: 39

Asn Ser Thr Trp Ala Pro Ser Phe Leu Ile Gln
 1 5 10

<210> SEQ ID NO 40
 <211> LENGTH: 11
 <212> TYPE: PRT
 <213> ORGANISM: *Streptococcus salivarius*

<400> SEQUENCE: 40

Lys Ser Thr Trp Ala Pro Ser Phe Leu Ile Lys
 1 5 10

What is claimed is:

1. A process of producing a fructo-oligosaccharide or fructo-polysaccharide, having $\beta(2-1)$ linked D-fructosyl units comprising forming a mixture by combining sucrose with at least one reaction partner selected from the group consisting of:

- a) a protein having fructosyltransferase activity, which exhibits at least 85% amino acid identity, as determined by a BLAST algorithm, with an amino acid sequence of SEQ ID No. 1, and
- b) a recombinant host cell containing one or more copies of a nucleic acid construct encoding for said protein (a) and capable of expressing said protein;

wherein said reaction partner interacts with sucrose to produce a fructo-oligosaccharide or fructo-polysaccharide.

2. The process according to claim 1, wherein said protein is a recombinant protein.

3. A process according to claim 1, further comprising chemically modifying said fructo-oligosaccharide or fructo-polysaccharide by simultaneous 3- and 4-oxidation, 1-or 6-oxidation, phosphorylation, acylation, hydroxyalkylation, carboxymethylation or amino-alkylation of one or more anhydrofructose units, or by hydrolysis.

4. The process according to claim 1, further comprising adding a food or beverage composition to said mixture to obtain a prebiotic composition.

5. The process according to claim 1, further comprising adding to said mixture a *Lactobacillus* strain capable of producing an oligosaccharide or polysaccharide and optionally a food or beverage composition, to obtain a synbiotic composition.

6. A process of producing a fructo-oligosaccharide or fructo-polysaccharide, having $\beta(2-1)$ linked D-fructosyl units comprising combining sucrose and a protein to form a mixture, said protein having fructosyltransferase activity, which exhibits at least 85% amino acid identity, as determined by a BLAST algorithm, with an amino acid sequence of SEQ ID No. 1, and

interacting said sucrose with said protein to produce said fructo-oligosaccharide or fructo-polysaccharide.

7. A process for producing a fructo-oligosaccharide or fructo-polysaccharide, having $\beta(2-6)$ linked D-fructosyl units comprising forming a mixture by combining sucrose with a reaction partner, wherein said reaction partner is a recombinant host cell containing one or more copies of a nucleic acid construct encoding for a protein having fructosyltransferase activity, which exhibits at least 85% amino acid identity, as determined by a BLAST algorithm, with an amino acid sequence of SEQ ID No. 11, and wherein said reaction partner interacts with sucrose to provide a fructo-oligosaccharide or fructo-polysaccharide.

8. A process according to claim 7, further comprising chemically modifying said fructo-oligosaccharide or fructo-polysaccharide by simultaneous 3- and 4-oxidation, 1-or 6-oxidation, phosphorylation, acylation, hydroxyalkylation, carboxymethylation or amino-alkylation of one or more anhydrofructose units, or by hydrolysis.

* * * * *

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER: _____**

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.